UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

THIAGO WILLIAN ALMEIDA BALSALOBRE

MAPEAMENTO DE QTLs EM POPULAÇÃO DERIVADA DE CRUZAMENTO COMERCIAL BI-PARENTAL EM CANA-DE-AÇÚCAR

QTL MAPPING IN POPULATION ORIGINATED FROM A BI-PARENTAL COMMERCIAL CROSS OF SUGARCANE

CAMPINAS
2016

THIAGO WILLIAN ALMEIDA BALSALOBRE

MAPEAMENTO DE QTLs EM POPULAÇÃO DERIVADA DE CRUZAMENTO COMERCIAL BI-PARENTAL EM CANA-DE-AÇÚCAR

QTL MAPPING IN POPULATION ORIGINATED FROM A BI-PARENTAL COMMERCIAL CROSS OF SUGARCANE

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção de título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology in the area of Plant Genetics and Genetic Breeding

Orientadora: PROF(A). DR(A). ANETE PEREIRA DE SOUZA
Co-orientadora: PROF(A). DR(A). MONALISA SAMPAIO CARNEIRO

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO THIAGO WILLIAN ALMEIDA BALSALOBRE E ORIENTADO DA PROF(A). DR(A). ANETE PEREIRA DE SOUZA

CAMPINAS
2016

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2010/50091-4 e 2008/52197-4

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Gustavo Lebre de Marco - CRB 8/7977

B216m

Balsalobre, Thiago Willian Almeida, 1986-
Mapeamento de QTLs em população derivada de cruzamento comercial bi-parental em cana-de-açúcar / Thiago Willian Almeida Balsalobre. – Campinas, SP : [s.n.], 2016.

Orientador: Anete Pereira de Souza.
Coorientador: Monalisa Sampaio Carneiro.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Cana-de-açúcar. 2. Marcadores moleculares. 3. Sequenciamento de DNA. 4. Mapeamento cromossômico. 5. Mapeamento de QTL. I. Souza, Anete Pereira de. II. Carneiro, Monalisa Sampaio. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: QTL mapping in population originated from a bi-parental commercial cross of sugarcane
Palavras-chave em inglês:
Sugar-cane
Molecular markers
DNA sequencing
Chromosome mapping
QTL mapping
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Anete Pereira de Souza [Orientador]
Renato Vicentini dos Santos
Marcelo Menossi Teixeira
Roberto Giacomini Chapola
João Ricardo Bachega Feijó Rosa
Data de defesa: 29-11-2016
Programa de Pós-Graduação: Genética e Biologia Molecular

Banca examinadora

Profa. Dra. Anete Pereira de Souza (orientadora)

Prof. Dr. Renato Vicentini dos Santos

Dr. Roberto Giacomini Chapola

Prof. Dr. Marcelo Menossi Teixeira

Dr. João Ricardo Bachega Feijó Rosa

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

Aos meus pais, Airton e Odete, à minha esposa, Fernanda, e a minha irmã, Mirelle

Dedico

Agradecimentos

Aos meus pais, Airton e Odete, por todo empenho na educação e criação dos filhos, pelos ensinamentos que mesmo sem palavras escritas ou faladas eram passados através das atitudes honestas e respeitosas, pelos incentivos, pela perseverança diante de situações adversas e pela coragem e alegria que sempre demonstraram. Sempre vocês serão exemplos de caráter e pessoas de bem. Esta tese de doutorado é, sem dúvida, também de vocês.

À minha esposa, Fernanda, pelos anos de convivência e de trabalho conjunto, desde nossas iniciações científicas, que acabou por gerar o amor mais verdadeiro o qual poderia sentir. Pela paciência que teve nos momentos difíceis, pela dedicação, carinho, amor, companheirismo e respeito, e por sempre estar ao meu lado zelando para que o melhor acontecesse em nossas vidas. Esta tese tem muito de você, da primeira letra ao último ponto final, sou grato pela força e coragem que me transmitiu. Obrigado!

À minha irmã, Mirelle, obrigado por entender minha ausência e cuidar dos nossos pais. Admiro seu amadurecimento, força e determinação.

À toda minha família e também a família que me acolheu (Zatti Barreto) pela preocupação, apoio e incentivo.

À minha co-orientadora Dra. Monalisa Sampaio Carneiro, que proporcionou crescimento profissional e pessoal, colaborou imensamente no desenvolvimento desta tese e sempre acreditou na minha capacidade e responsabilidade para fazer ciência. Sou imensamente grato por toda dedicação, entusiasmo, paciência e pelas conversas que direcionavam e mostravam caminhos melhores para seguir. Esta tese não existiria sem sua imensa colaboração.

À minha orientadora Dra. Anete Pereira de Souza, pela oportunidade de seguir seus ensinamentos, compartilhar as recentes inovações científicas para o complexo genoma da cana-de-açúcar e de realizar o Doutorado na UNICAMP em um programa de pós-graduação reconhecido pela qualidade dos trabalhos desenvolvidos.

A todos os colegas do Laboratório de Genética Molecular (LAGEM) da UFSCar, do Laboratório de Biotecnologia de Plantas (LBP) da UFSCar e do Laboratório de Análise

Genética e Molecular (LAGM) da UNICAMP, pela oportunidade de crescimento conjunto e por todos os momentos compartilhados.

Ao professor Dr. Alfredo Seiti Urashima, coordenador do LAGEM, pelos anos de convivência e amizade.

Aos técnicos do LAGEM, LAGM e LBP, pela ajuda constante durante todo o período de desenvolvimento desta tese. Em especial à Liliane Trento Scorzoni, Daniele Rebelatto e Isabella Barros Valadão do LBP.

Ao laboratório do Professor Dr. Antônio Augusto Franco Garcia (ESALQ-USP), por auxiliar em todas as análises genético-estatísticas e especialmente a ele pelos ensinamentos e tratamento sempre gentil, e ao colega Guilherme da Silva Pereira, pela ajuda com as análises dos dados fenotípicos e genéticos, e também por compartilhar os ensinamentos que recebeu durante o período do seu doutorado.

Ao Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA) da UFSCar pertencente à RIDESA (Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético) pela criação e manutenção dos experimentos utilizados nesta tese e por todo o suporte técnico necessário para as avaliações fenotípicas. Em especial ao coordenador do PMGCA da UFSCar, Dr. Hermann Paulo Hoffmann, por disponibilizar todo este suporte, aos técnicos Luiz Plínio Zavaglia e Sandro Augusto Ferrarez, pelo cuidado, empenho e dedicação no auxílio da condução dos experimentos de campo e também ao Dr. Rodrigo Gazaffi, pelo auxílio com análises no software R e mapeamento de QTLs.

Aos membros da pré-banca e aos da banca de defesa, pela disponibilidade em contribuir para que esse trabalho fosse melhorado.

À FAPESP, pela concessão da bolsa de doutorado (Processo n°: 2010/50091-4) e apoio financeiro para a realização desta tese (Processo n°: 2008/52197-4).

Resumo

A cana-de-açúcar é uma fonte renovável de energia e com potencial para expansão. A complexidade genética da cana-de-açúcar decorrente de seu alto nível de ploidia e aneuploidia, aliada à natureza quantitativa da maioria dos caracteres agronômicos tem dificultado, atualmente, a obtenção de elevados índices de ganho genético através do melhoramento convencional. O desenvolvimento de marcadores moleculares e a construção de mapas genéticos podem auxiliar na elaboração de estratégias a serem introduzidas nos programas de melhoramento de forma a aumentar a eficiência dos processos de seleção e acelerar o desenvolvimento de novas cultivares. Desta forma, a proposta desta tese foi a construção de um mapa genético em cana-de-açúcar visando a identificação de regiões genômicas que controlam características de interesse através do mapeamento de QTL (*Quantitative Trait Loci*). Uma população de cana-de-açúcar com 153 indivíduos oriundos do cruzamento comercial bi-parental entre as cultivares SP80-3280 e RB835486 foi utilizada para alcançar os objetivos citados. O experimento de campo foi instalado em duas localidades, Araras-SP e Ipaussu-SP, usando o delineamento de blocos aumentados incompletos com 3 repetições. As avaliações fenotípicas foram realizadas ao longo de três anos (2011, 2012 e 2013). Empregou-se a abordagem de modelos mistos para análise das características fenotípicas relacionadas com componentes de produção e resistência à ferrugem marrom. Os dados de severidade à ferrugem marrom foram analisados, como uma primeira abordagem, via modelo misto linear generalizado. As estimativas de herdabilidade das características fenotípicas foram altas, variando de 0.78 (altura de colmos) a 0.92 (diâmetro de colmos), e a análise de severidade à ferrugem marrom mostrou que 66% dos clones possuem, no mínimo, 90% de probabilidade de serem resistentes à doença. Para construção do mapa genético integrado foram utilizados marcadores moleculares do tipo microssatélites (*Simple Sequence Repeats*, SSR), TRAP (*Target Target Region Amplification Polymorphism*), além de SNPs (*Single Nucleotide Polymorphims*) e indels (inserções e deleções) oriundos da técnica de GBS (*Genotyping-by-Sequencing*). Para descoberta de marcadores baseados em GBS foram utilizadas quatro pseudo-referências: genoma de sorgo (*Sorghum bicolor*), genoma metil-filtrado da cana-de-açúcar, transcriptoma da cana-de-açúcar (RNAseq) e sequências do projeto SUCEST. A ploidia e dosagem de cada loco bi-alélico foi estimada através do software SUPERMASSA. Utilizando o software Onemap (v. 2.0-4) e empregando-se $LOD > 9.0$ e fração de recombinação < 0.10 foram mapeados 993 marcadores em dose única. Estes foram agrupados em 223 grupos de ligação e 18 grupos de homo(eo)logia. A extensão total do

mapa foi 3,682.04 cM e a densidade de marcadores foi de 3.70 cM. Utilizando mapeamento por intervalo composto (*Composite Interval Mapping*, CIM) foram mapeados sete QTLs considerando quatro das 11 características fenotípicas avaliadas. Os resultados sugerem a presença de um QTL estável entre locais para conteúdo de sólidos solúveis (BRIX) e para teor de sacarose (POL%C). Além disso, QTLs para BRIX e teor de fibra (FIB) tiveram marcadores associados com genes candidatos com grande potencial de validação e consequente uso no melhoramento molecular de cana-de-açúcar. Este estudo é o primeiro a reportar o uso de GBS para descoberta de variantes em larga escala e genotipagem de uma população de cana-de-açúcar com posterior análise de mapeamento por intervalo composto.

Palavras-chave: *Saccharum* spp., marcadores moleculares, GBS, mapa genético, seleção assistida por marcador

Abstract

Sugarcane is a renewable source of energy and with potential for expansion. The genetic complexity of sugarcane due to its high level of ploidy and aneuploidy, together with the quantitative nature of most agronomic traits, has hindered currently the achievement of high rates of genetic gain through conventional breeding of this crop. The development of molecular markers and construction of genetic maps can be helpful to establish strategies in breeding programs and increase the efficiency of the selection process and accelerate the development of new cultivars. Therefore, the purpose of this thesis was to construct an integrated genetic map in sugarcane and identify genomic regions that control traits of interest by QTL mapping (*Quantitative Trait Loci*). To achieve these goals we used an F1 segregating population of sugarcane with 153 individuals from bi-parental commercial cross between cultivars SP80-3280 and RB835486. The field trial was carried out in two locations, Araras-SP and Ipaussu-SP, and the experimental design consisted of an augmented randomized incomplete block, which was fully replicated three times. The phenotypic evaluations were performed over three years (2011, 2012 and 2013). We applied a mixed model approach for the analysis of phenotypic traits related to yield components and resistance to brown rust. The severity data of brown rust was analyzed by generalized linear mixed model. Heritability estimates were high, ranging from 0.78 (stalk height) to 0.92 (stalk diameter), and the brown rust severity analysis showed that 66% of the clones have at least 90% probability of being resistant to disease. To construct an integrated genetic map were used molecular markers microsatellites (*Simple Sequence Repeats*, SSR), TRAP (*Target Target Region Amplification Polymorphism*), as well as SNPs (*Single Nucleotide Polymorphims*) and indels derived from the GBS protocol (*Genotyping-by-Sequencing*). To GBS-based markers discovery were used four pseudo-references: sorghum genome (*Sorghum bicolor*), methyl-filtered genome of sugarcane, transcriptome of sugarcane (RNAseq) and sequences of SUCEST project. The ploidy and allelic dosage of each bi-allelic locus was estimated by SUPERMASSA software. Using Onemap software (v. 2.0-4) and employing LOD> 9.0 and recombination fraction <0.10, a total of 993 markers in single dose were mapped. These markers were distributed throughout 223 linkage groups, which were clustered in 18 homo(eo)logous groups. The total length of the map was 3,682.04 cM with an average marker density of 3.70 cM. Using composite interval mapping (*Composite Interval Mapping*, CIM) were mapped seven QTLs considering four of the 11 phenotypic traits evaluated. The results suggest the presence of a stable QTL across locations to soluble solid content (BRIX) and sucrose content of the cane

(POL%C). Furthermore, QTLs to BRIX and fiber content (FIB) traits had associated markers with candidate genes, which had great potential for validation and future use for molecular breeding in sugarcane. This study is the first to report the use of GBS for large-scale variant discovery and genotyping of a population in sugarcane with posterior analysis to composite interval mapping.

Keywords: *Saccharum* spp., molecular markers, GBS, genetic map, QTL, marker-assited selection

Sumário

A cana-de-açúcar tem grande importância visto que movimenta diversos setores da economia brasileira e mundial com a produção principalmente de açúcar, bioetanol e energia elétrica a partir do bagaço. Somente no Brasil, a movimentação financeira do setor sucroenergético pode ultrapassar 22 bilhões de dólares por safra. A cana-de-açúcar moderna é predominantemente autopoliploide e possui o genoma mais complexo que qualquer outra cultura melhorada apresentando variável nível de ploidia e frequente aneuploida. Desta forma, o objetivo dessa tese foi contribuir para uma maior compreensão da estrutura genética da cana-de-açúcar através da construção de um mapa genético e do mapeamento de QTL (*Quantitative Trait Loci*) de características de importância econômica. Uma população originada de um cruzamento comercial bi-parental, realizado pelo Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA) da Universidade Federal de São Carlos (UFSCar), foi utilizada para esta finalidade. Os resultados obtidos nesta pesquisa estão apresentados em dois capítulos no formato de artigos. O primeiro artigo, com título "*Mixed Modeling of Yield Components and Brown Rust Resistance in Sugarcane Families"* foi publicado no periódico *Agronomy Journal* (108:1–14 (2016) / doi:10.2134/agronj2015.0430). Já o segundo artigo, com título "*GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane"* foi aceito (10/10/2016) no periódico *BMC Genomics.*

O Capítulo 1 descreve as avaliações fenotípicas de 11 características de importância econômica em duas famílias de cana-de-açúcar. As análises dos dados obtidos para estimativa dos parâmetros genéticos foram realizadas através da abordagem de modelos mistos. Neste capítulo são apresentados resultados para duas famílias nomeadas de SR1 e SR2. A primeira família possui 153 indivíduos oriundos do cruzamento bi-parental entre as cultivares SP80-3280 e RB835486, cujos resultados fazem parte do meu projeto de doutorado. Já a segunda família possui 240 indivíduos oriundos do cruzamento bi-parental entre as cultivares SP81-3250 e RB925345, cujos resultados fazem parte do doutorado, já defendido, pela Dra. Melina Cristina Mancini, integrante do mesmo grupo de pesquisa. Os experimentos de campo foram conduzidos em duas localidades, Araras-SP e Ipaussu-SP, e com três repetições. As avaliações fenotípicas foram realizadas ao longo de três anos (2011, 2012 e 2013) para as seguintes características: número de perfilhos (*stalk number*, SN), altura de colmos (*stalk height*, SH, em cm), diâmetro de colmos (*stalk diameter*, SD, em mm), peso de colmos da parcela (*stalk weight*, SW, em Kg), produção estimada em toneladas de cana por hectare (*cane yield*, TCH), teor de sólidos solúveis (*soluble solid content*, BRIX, em ºBrix), teor de fibra da cana (*fiber content*, FIB, em porcentagem), teor de sacarose de caldo

(*sucrose content of juice*, POL%J, em porcentagem), teor de sacarose da cana (*sucrose content of cane*, POL%C, em porcentagem), rendimento estimado em toneladas de sacarose por hectare (*sucrose yield*, TPH) e resistência à ferrugem marrom que foi analisada via modelo misto linear generalizado. Avaliando a família SR1, as estimativas de herdabilidade foram altas, variando de 0.78 (SH) a 0.92 (SD), e a análise de severidade à ferrugem marrom mostrou que 66% dos clones possuem, no mínimo, 90% de probabilidade de serem resistentes a essa doença. Em geral, as estimativas dos parâmetros genéticos obtidos através de modelos mistos refletiram resultados próximos aos já relatados para cultura, no entanto, houve uma melhora na interpretação dos resultados. Isto porque tais parâmetros genéticos foram obtidos a partir das estruturas mais prováveis de variâncias e covariâncias genéticas para as matrizes de locais e colheitas avaliadas para cada característica. De forma geral, estruturas que consideram heterogeneidade de variâncias e presença de correlações genéticas foram selecionadas, diferindo-se dos modelos de análise tradicionais.

O Capítulo 2 apresenta a construção de um mapa genético em cana-de-açúcar com marcadores microssatélites (*Simple Sequence Repeats*, SSR), TRAP (*Target Target Region Amplification Polymorphism*), e SNPs (*Single Nucleotide Polymorphims*) e indels (inserções e deleções) oriundos da técnica de GBS (*Genotyping-by-Sequencing*), além do mapeamento de QTLs através do modelo CIM (*Composite Interval Mapping*) para quatro das 11 características fenotípicas que foram apresentadas no Capítulo 1. Para descoberta de SNPs e indels foram utilizadas quatro pseudo-referências: genoma de sorgo (*Sorghum bicolor*), genoma metil-filtrado da cana-de-açúcar, transcriptoma da cana-de-açúcar (RNAseq) e sequências do projeto SUCEST. A ploidia e dosagem de cada loco SNP foi estimada através do software SUPERMASSA (Serang *et al*., 2012). O mapa genético foi construído utilizando o software Onemap (v. 2.0-4) (Margarido *et al*., 2007) e empregando LOD > 9.0 e fração de recombinação < 0.10. Do total de 7,678 marcadores em dose única considerados para a análise de ligação, 993 foram mapeados, incluindo marcadores com segregação 1:1, 1:2:1 e 3:1. Os marcadores mapeados formaram 223 grupos de ligação e 18 grupos de homo(eo)logia com cobertura total de 3,682.04 cM. A densidade média de marcadores foi de 3.70 cM. A análise de mapeamento resultou em sete QTLs detectados, sendo três para BRIX, dois para POL%C, um para SD e um para FIB. Os resultados sugerem a presença de um QTL comum e estável entre locais para BRIX e POL%C. Além disso, QTLs para BRIX e FIB tiveram marcadores associados com genes candidatos com grande potencial de validação e consequente uso no melhoramento molecular de cana-de-açúcar.

Um estudo paralelo foi realizado utilizando o sequenciamento de nova geração (*Next Generation Sequencing*, NGS) em equipamento da Illumina para obtenção do transcriptoma de seis cultivares de cana-de-açúcar, que deram origens a três populações distintas de mapeamento genético (IACSP96-3046 e IACSP95-3018, SP81-3250 e RB925345, e SP80-3280 e RB835486). Deste trabalho resultou o artigo científico "*De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties*" (Anexo I) publicado no periódico *PLoS ONE* (doi: 10.1371/journal.pone.0088462). Com este estudo foi possível obter genes únicos para a cana-de-açúcar e identificar um grande número de marcadores moleculares (SSR e SNPs). Até a data de finalização desta tese, o artigo já possuía 21 citações (*Google Scholar*), demonstrando a relevância do estudo tanto pela metodologia utilizada quanto pelos resultados obtidos.

Também como estudo paralelo, três novas cultivares de cana-de-açúcar liberadas comercialmente pelo PMGCA da UFSCar, integrante da RIDESA (Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético), foram caracterizadas quanto à época de maturação, produtividade e *fingerprinting* com marcadores SSR. Duas destas, RB965902 e RB965917, estão descritas no artigo "*RB965902 and RB965917 - Early/medium maturing sugarcane varieties*", enquanto que a terceira cultivar, RB975952, está descrita no artigo "*RB975952 - Early maturing sugarcane cultivar*". Ambos os artigos foram publicados no periódico *Crop Breeding and Applied Biotechnology* em 2011 e 2014, respectivamente (Anexo II). Outras duas cultivares liberadas comercialmente em 2015, RB975242 e RB975201, também foram caracterizadas e o manuscrito intitulado *"RB975242 and RB975201 - Late maturation sugarcane varieties"* foi aceito para publicação no periódico *Crop Breeding and Applied Biotechnology*. De acordo com o censo varietal de 2016, realizado pelo PMGCA da UFSCar, levando em consideração 124 unidades produtoras de São Paulo e Mato Grosso do Sul, dois dos Estados produtores de cana-de-açúcar do país, as cultivares RB975201 e RB965902 ocupam a oitava e décima colocação, respectivamente, dentre as mais plantadas.

Esta tese conta ainda com um tópico de resultados complementares, uma discussão geral, um resumo dos resultados obtidos e uma conclusão geral. O estudo genético da cana-de-açúcar é um constante desafio visto que ainda caminhamos para encontrar as melhores metodologias de obtenção e de análises de dados genômicos. A complexidade do genoma da cana não é maior que o entusiasmo por novas descobertas e, através da aplicação de técnicas da biologia molecular e da bioinformática, é razoável pensar que maiores índices de produtividade poderão ser obtidos, como já ocorre para outras culturas, como milho e soja.

Uma visão geral sobre a cultura da cana-de-açúcar e os marcadores moleculares

A matriz energética do Brasil é composta por 43,5% de fontes renováveis, sendo os produtos derivados da cana-de-açúcar responsáveis por aproximadamente 42% deste total (EPE, 2015). Além disso, o Brasil é o maior produtor de cana-de-açúcar do mundo, com 737 milhões de toneladas de cana colhidas em 2014 em aproximadamente nove milhões de hectares (FAOSTAT, 2014; CONAB, 2016).

O interesse econômico na cultura da cana-de-açúcar tem aumentado nos últimos anos devido à alta demanda mundial por energia sustentável (Cheavegatti-Gianotto *et al*., 2011). Como consequência do seu potencial energético, a cana-de-açúcar vem se expandindo principalmente em ambientes tropicais e subtropicais (Singh *et al*., 2010; Aitken *et al*., 2014). Além do açúcar e do etanol, a agroindústria da cana-de-açúcar apresenta grandes possibilidades de diversificação de seus produtos, seja caminhando em direção às biorrefinarias, complexos produtivos capazes de fornecer bioenergia e fabricar biomateriais diversos (como exemplo a produção de plásticos biodegradáveis a partir do bagaço da cana-de-açúcar), seja reforçando a base de unidades mais antigas com adoção de novas tecnologias para co-geração de energia elétrica ou reaproveitamento de leveduras, que foram utilizadas no processo fermentativo, para extração de proteínas de alto valor comercial. Em adição, o próprio aumento da utilização da bioenergia e do etanol como biocombustível gera a necessidade de aumentar a produtividade e expandir a área de cultivo da cana.

A produtividade média da cana-de-açúcar no Brasil é de, aproximadamente, 72 toneladas por hectare, entretanto o potencial de produção teórico é de 380 toneladas por hectare (Waclawovsky *et al*., 2010; CONAB, 2016). A utilização de práticas agrícolas adequadas como irrigação, fertilização, controle de pragas e colheita em época de maturação máxima pode proporcionar maiores rendimentos e beneficiar toda a cadeia bioenergética da cana-de-açúcar. Além disso, o melhoramento pode obter ganhos genéticos que irão impactar diretamente sobre a produtividade. Entre as características agronômicas de interesse comercial, que são alvo do melhoramento genético, encontramos teor de sacarose, produção em toneladas de cana por hectare, quantidade e qualidade de fibra, resistência a doenças e tolerância à seca.

Os programas de melhoramento genético da cana-de-açúcar objetivam o aumento da produção e da produtividade da cana através da seleção de genótipos superiores com base

em populações segregantes obtidas a partir do cruzamento entre certas cultivares, em um trabalho de longa duração. Recentemente, foram observados menores aumentos na produção de açúcar, cerca de 1 a 1,5% ao ano (Dal-Bianco *et al*., 2012; CONAB, 2016). Nesse contexto, o desenvolvimento e lançamento de novas cultivares com características agronômicas que atendam às demandas atuais e as perspectivas do setor sucroenergético se deparam com características genéticas complexas e genuínas da cana-de-açúcar. Pertencente à família das gramíneas (Poaceae), tribo Andropogoneae, a cana-de-açúcar faz parte de um complexo de espécies poliplóides do gênero *Saccharum*, no qual ocorrem seis espécies: *S. officinarum* L. (2n = 80), *S. robustum* J. (2n = 60-205), *S. barberi* J. (2n = 81-124), *S. sinense* R. (2n = 111-120), *S. spontaneum* L. (2n = 40-128) e *S. edule* H. (2n = 60-80) (Daniels & Roach,1987). O elevado nível de ploidia (2n = 100-130), a ocorrência de aneuploidia e a complexidade citogenética dos híbridos modernos interespecíficos (Figura 1) (D' Hont *et al*., 1996; D'Hont *et al*., 1998; D'Hont & Glaszmann, 2001; D'Hont, 2005; Piperidis *et al*., 2010), obtidos principalmente pelo cruzamento entre *S. officinarum* L. (2n = 80), rico em teor de açúcar, e *S. spontaneum* L. (2n = 40-128), altamente tolerante a estresses bióticos, aliado ao tempo gasto com o desenvolvimento de novas cultivares (de 10 a 15 anos), são barreiras que devem ser consideradas para obtermos maiores ganhos genéticos em cana-de-açúcar. Portanto, para atender os anseios de aumento da produtividade é necessário aprimorar as técnicas do melhoramento genético.

O desenvolvimento de cultivares superiores pelo melhoramento convencional poderá ser mais eficiente se puder ser assistido por marcadores moleculares, os quais possibilitam a identificação de regiões genômicas que controlam características de interesse comercial e fornecem estimativas mais confiáveis de diversidade genética por serem independentes de efeitos ambientais. Este conhecimento é extremamente importante para a orientação dos cruzamentos nos programas de melhoramento genético, já que a escolha eficiente dos genitores é um passo fundamental para obtenção de cultivares mais produtivas.

Diversos tipos de marcadores moleculares estão disponíveis para análise do genoma de cana-de-açúcar, dentre eles, temos os marcadores TRAP (*Target Target Region Amplification Polymorphism*), observados através da amplificação de região flanqueada por um *primer* arbitrário e outro fixo em região gênica, e os marcadores microssatélites (*Simple Sequence Repeats*, SSR), observados através da amplificação de regiões com repetições em tandem de pequenas sequências de nucleotídeos. Tais marcadores têm sido utilizados para estudos básicos de genética em cana-de-açúcar e no processo de melhoramento, incluindo caracterização molecular de germoplasma, identificação varietal, avaliação do grau de

parentesco e mapeamento genético (Cordeiro *et al*., 2003; McIntyre *et al*., 2005; Garcia *et al*., 2006; Pinto *et al*., 2004, 2006; Alwala *et al*., 2006; Raboin *et al*., 2006; Creste *et al*., 2010; Palhares *et al*., 2012; Pastina *et al*., 2012).

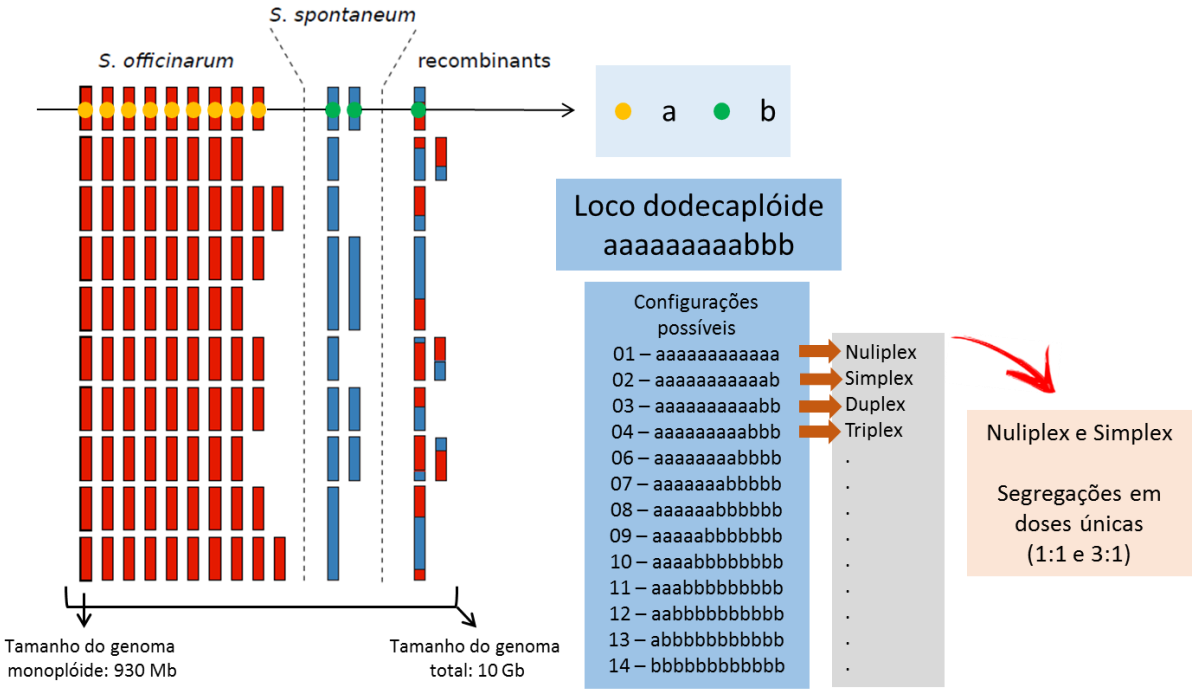


Figura 1 - Organização cromossômica, adaptado de D'Hont, (2005), do híbrido moderno interespecífico R570. Esta cultivar possui 2n = 115 e tamanho total do genoma de aproximadamente 10 gigabases (Gb). Cerca de 80% dos cromossomos de R570 foram originados da espécie *Saccharum officinarum* (barras vermelhas), cerca de 10% foram originados da espécie *S. spontaneum* (barras azuis) e cerca de 10% foram recombinantes (barras vermelhas e azuis). As bolinhas de cores amarelas e verdes representam alelos '*a*' e '*b*', respectivamente, em um loco dodecaploide tomado ao acaso. Este loco está inserido em um dos 10 conjuntos cromossômicos representados; baseados nos cromossomos oriundos de *S. officinarum*. O nível de ploidia dentro de cada conjunto cromossômico é variável. De todas as configurações possíveis de segregação, apenas segregações em doses únicas (1:1, originada a partir do cruzamento entre simplex e nuliplex; e 3:1 ou 1:2:1, originada a partir do cruzamento entre simplex e simplex) são aproveitadas atualmente para estudos de mapeamento genético em cana-de-açúcar.

Com base em informações de projeto de transcriptoma, diversos trabalhos identificaram a presença de locos SSR derivados de sequências expressas (Vettore *et al*., 2003; Oliveira *et al*., 2009; Marconi *et al*., 2011; Cardoso-Silva *et al*., 2014), revelando a importância destes marcadores nas investigações genéticas a nível funcional do genoma de cana (Pinto *et al*., 2006; Oliveira *et al*., 2007). Recentemente, a genotipagem com os marcadores SNPs (*Single Nucleotide Polymorphisms*), observados através da alteração de um

único nucleotídeo na sequência do genoma, pode representar um grande avanço na análise genética da cana-de-açúcar por serem encontrados em grande número e por permitirem a estimação da dosagem alélica (Henry *et al.*, 2012; Serang *et al*., 2012; Garcia *et al*., 2013; Costa *et al.*, 2016). As novas abordagens para genotipagem baseadas em sequenciamento de nova geração (*Next Generation Sequencing*, NGS), como o protocolo de genotipagem por sequenciamento (*Genotyping-by-Sequencing*, GBS), permite a detecção de centenas de milhares de SNPs, reduzindo custos no estudo da genômica de organismos complexos como a cana-de-açúcar (Elshire *et al*., 2011; Glaubtiz *et al*., 2014).

Desta forma, com o emprego de marcadores moleculares, baseado em ferramentas da engenharia genética e biologia molecular, e com as novas metodologias de análise de dados genômicos, o melhoramento genético convencional da cana-de-açúcar poderá aumentar a eficiência de todo o processo de obtenção de cultivares mais produtivas, sustentando um dos pilares para a expansão do setor sucroenergético.

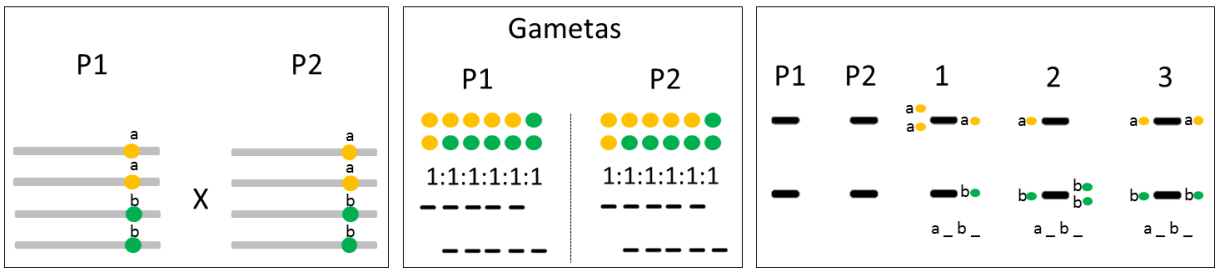
Mapeamento de QTLs em cana-de-açúcar

O termo *Quantitative Trait Loci*, ou QTL, tem sido comumente utilizado para denominar regiões cromossômicas que contêm genes (ou locos) que controlam caracteres poligênicos, isto é, caracteres quantitativos, com padrão contínuo de variação fenotípica (Falconer & Mackay, 1996; Liu, 1998; Lynch & Walsh, 1998). A partir das informações fornecidas por marcadores moleculares, é possível mapear locos que controlam tais caracteres (Ferreira & Grattapaglia, 1998). Mapear QTLs significa fazer inferências em todo o genoma sobre as relações entre o genótipo e o fenótipo de caracteres quantitativos. Essas inferências incluem informações sobre número, posição, efeitos, interações dos alelos dos QTLs dentro dos locos (dominância) e entre os locos (epistasia), efeitos pleiotrópicos dos QTLs e interações entre QTLs e ambientes (Jiang & Zeng, 1995; Zeng, Kao & Basten, 1999; Zeng, 2001; Malosetti *et al*., 2004; Malosetti *et al*., 2008). Para tanto, é necessária uma população que apresente variabilidade genética e elevado desequilíbrio de ligação. Além disso, são necessárias metodologias genético-estatísticas sofisticadas com forte suporte computacional, devido à complexidade das análises tanto fenotípicas como genotípicas (Margarido, 2011). Estudos de mapeamento de QTLs são usualmente realizados em duas etapas. Primeiro, as médias fenotípicas ajustadas são obtidas; segundo, essas médias são utilizadas para buscar associações com marcadores moleculares e/ou ao longo de mapas genéticos.

A cana-de-açúcar é uma espécie semi-perene, em que vários cortes são conduzidos, de modo que os indivíduos ficam sujeitos a diferentes condições ambientais ao longo dos anos, e é importante identificar QTLs que se expressem de forma consistente ao longo de cortes e/ou locais ou que se expressem sob condições específicas de um determinado ambiente. Em programas de melhoramento genético, é comum a avaliação de diversos indivíduos em diferentes ambientes, nos chamados *Multi-Environment Trials* (METs), bem como a avaliação de diversos caracteres simultaneamente, visto que genótipos superiores devem concentrar alelos favoráveis para produção, resistência a doenças, pragas e estresses abióticos, caracteres agronômicos, entre outros (Welham *et al*., 2010). Com o intuito de analisar estes dados fenotípicos, os modelos mistos exibem vantagens que os tornam superiores aos modelos lineares tradicionais. Entre elas, destacam-se: suposições mais realistas com relação à estrutura de correlação entre os resíduos, levando-se em consideração fontes sistemáticas e aleatórias de erro no campo experimental, facilidade de lidar com dados incompletos e desbalanceados e a possibilidade de considerar alguns efeitos como aleatórios ao invés de fixos (Smith, Cullis & Thompson, 2005). O uso de modelos mistos também é interessante por melhorar a parte experimental da análise, visto que fontes de variação adicionais, como o efeito de blocos e de covariáveis são incluídas diretamente na análise, em contraste ao simples uso de médias aritméticas como é feito usualmente (Pastina, 2010; Margarido, 2011; Pastina *et al.*, 2012).

Vários métodos para mapear QTLs já foram propostos, dentre eles a análise por marcas individuais (Weller, 1986; Lynch & Walsh, 1998; Edwards, Stuber & Wendel, 1987), Mapeamento por Intervalo (*Interval Mapping*, IM) (Lander & Botstein, 1989; Lynch & Walsh, 1998), Mapeamento por Intervalo Composto (*Composite Interval Mapping*, CIM) (Zeng, 1993; Zeng, 1994; Jansen & Stam, 1994) e Mapeamento por Múltiplos Intervalos (*Multiple Interval Mapping*, MIM) (Kao & Zeng, 1997; Kao, Zeng & Teasdale, 1999). Recentemente, Gazaffi *et al*. (2014), desenvolveram um modelo para mapear QTLs em populações biparentais de irmãos completos (F1 segregante) em espécies que dificilmente conseguem obter linhas puras e que sofrem grande depressão por endogamia devido às autofecundações, como é o caso da cana-de-açúcar. Nessas populações as fases de ligação entre diferentes locos e QTLs não são conhecidas *a priori*, o que dificulta as análises de ligação e o mapeamento de QTLs. Assim o modelo desenvolvido, expandido de Lin *et al*. (2003) para o contexto do CIM (Zeng, 1994; Jiang & Zeng, 1997) também possibilita identificar a fase de ligação e a localização do QTL a partir de marcadores com diferentes segregações nos genitores.

Os primeiros trabalhos de associação entre dados fenotípicos e marcadores moleculares em cana-de-açúcar surgiram simultaneamente à publicação dos primeiros mapas de ligação e quase sempre envolveram cruzamentos biparentais, interespecíficos ou não. Em geral, tais trabalhos de mapeamento genético em cana-de-açúcar resultaram na geração de dois mapas, um para cada genitor (*pseudo-testcross*) (Grattapaglia & Sederoff, 1994), com base em marcadores segregando na proporção de 1:1 (Wu *et al*., 1992). Além disso, diversos trabalhos de mapeamento de QTLs foram desenvolvidos para essa espécie através de análises de marcas individuais (Sills *et al*., 1995; Daugrois *et al*., 1996; Guimarães, Sills & Sobral, 1997; Asnaghi *et al*., 2001; Ming *et al*., 2001, Ming *et al*., 2002a, 2002b; Hoarau *et al*., 2002; Jordan *et al*., 2004; Silva & Bressiani, 2005; McIntyre *et al*., 2005a, 2005b, 2006; Reffay *et al*., 2005; Aitken, Jackson & McIntyre, 2006; Raboin *et al*., 2006, 2008; Wei *et al*., 2006; Al-Janabi *et al*., 2007; Oliveira *et al*., 2007; Aitken *et al*., 2008; Piperidis *et al*., 2008; Pinto *et al*., 2010). Algumas exceções existem, visto que alguns trabalhos incluíram marcadores com segregação 3:1 (Hoarau *et al*., 2002; McIntyre *et al*., 2005a, 2005b; Reffay *et al*., 2005; Aitken, Jackson & McIntyre, 2006; Raboin *et al*., 2006; Al-Janabi *et al*., 2007; Oliveira *et al*, 2007; Aitken *et al*., 2008; Piperidis *et al*., 2008; Pastina *et al*., 2012; Singh *et al*., 2013; Margarido *et al*., 2015) e 1:2:1 (Costa *et al*., 2016). Dosagens maiores que 1:1, 3:1 e 1:2:1 não são, até o momento, utilizadas em trabalhos de mapeamento em espécies poliploides devido, principalmente, a falta de modelos genético-estatíticos capazes de inserir marcadores em múltiplas doses nos mapas genéticos (Figura 2).



a

b

Figura 2 - Exemplo da utilização de marcadores moleculares baseados em gel, como microssatélites, em organismo autotetraploide, adaptado de Garcia & Mollinari (2013). **Fig**

2a: Configuração alélica dos genitores P1 (*aabb*) e P2 (*aabb*), gametas formados pelos dois genitores e simulação dos fragmentos amplificados pelo marcador molecular nos genitores e em uma progênie hipotética. Apesar dos fragmentos indicarem a presença dos alelos '*a*' (em amarelo) e '*b*' (em verde) na progênie 1, 2 e 3, não é possível determinar a dosagem desses alelos em cada um dos fragmentos observados. **Fig 2b:** Representação da associação entre a configuração alélica de um gene e o fenótipo teor de açúcar. Em poliploides, essa associação entre forma alélica e fenótipo ainda não é realizada. Os marcadores SNPs podem ter estimados a ploidia e a dosagem alélica através do software SUPERMASSA, o que representa um avanço sobre os marcadores moleculares baseados em gel, no entanto, ainda faltam ferramentas genético-estatísticas capazes de mapear os marcadores que estão em múltiplas doses.

Como exemplo de estudo de mapeamento de QTLs, Pastina *et al*. (2012) utilizando uma progênie derivada do cruzamento bi-parental entre as cultivares SP80-180 e SP80-4966, marcadores derivados de sequências expressas e dados fenotípicos de dois locais e três colheitas consecutivas, detectaram um total de 46 QTLs. Neste trabalho, foi utilizada uma estratégia com modelos mistos e mapeamento por intervalo para mapear QTLs de produção de cana (toneladas de cana por hectare), produção de açúcar (toneladas de açúcar por hectare), porcentagem de fibra e teor de sacarose. O mapeamento de QTLs em espécies poliplóides, como a cana-de-açúcar, ainda é muito importante devido à complexidade genética e as metodologias que ainda se desenvolvem para a análise dos dados fenotípicos e genéticos. Com o advento das novas plataformas de genotipagem baseadas em sequenciamento, as quais são capazes de identificar milhares de marcadores SNPs no genoma, e com o avanço da análise de dados em organismos poliploides (Serang *et al*., 2012; Garcia *et al*., 2013), a pesquisa da arquitetura genética da cana-de-açúcar pode avançar para descobertas que vislumbram uma aplicação prática no aumento de rendimentos diretamente no campo.

NGS e identificação de SNPs: uma nova estratégia para construção de mapas genéticos de alta densidade em cana-de-açúcar

Há uma tendência em utilizar marcadores SNPs para substituir outros tipos de marcadores em muitas espécies, uma vez que são frequentemente comuns no genoma, dentro e entre os genes (Bundock *et al*., 2009). Os SNPs são variações na sequência de DNA que ocorrem quando um único nucleotídeo é alterado. Tais variações representam polimorfismos que, juntamente com as deleções e inserções, são responsáveis pela maior parte da variação genética nos organismos (Cho *et al*., 1999; Rafalsky & Tingey, 2008) e são amplamente

distribuídos pelo genoma, sendo mais abundantes em regiões não transcritas e em regiões que flanqueiam os microssatélites (Mogg *et al*., 2002; Bundock & Henry, 2004).

A alta freqüência de SNPs em muitas espécies de plantas, tais como milho (Tenaillon *et al*., 2001), cevada (Kanazin *et al*., 2002) e arroz (Yu *et al*., 2002), torna esse marcador uma escolha eficiente para o diagnóstico de doenças, seleção assistida por marcadores (SAM), mapeamento genético de alta resolução e caracterização varietal (Batley *et al*., 2003) e mapeamento por associação. Além de apresetarem distribuição ampla, os SNPs permitem análises em escalas maiores quando comparado aos SSR (Garcia *et al*., 2013). Isso acontece em decorrência, dentre outros fatores, aos processos mais simples e rápidos que coletivamente facilitam a automação e o monitoramento da amostra, além de poderem proporcionar maior quantidade de dados (Jones *et al*., 2007).

Uma das estratégias para descoberta de SNPs é a utilização de bases de dados de ESTs (*Expressed Sequence Tags*) públicas. A busca de SNPs em bancos de sequências vem sendo realizada com sucesso em diferentes espécies que possuem grande quantidade de ESTs em bancos de dados (Buetow *et al*., 1999; Batley *et al*., 2003; Kota *et al*., 2003). As bases de dados de ESTs que foram utilizadas na identificação de SNPs para cana-de-açúcar até o momento foram o SUCEST (Vettore *et al*., 2003; Grivet *et al*., 2003; McIntyre *et al*., 2006; Garcia *et al*., 2013; Costa *et al*., 2016) e o Plantdb (Cordeiro *et al*., 2006). O desenvolvimento de SNPs identificados a partir de ESTs representa um valioso sistema de marcadores no mapeamento de genes candidatos e na identificação da base genética de QTLs de características de importância agronômica.

Os recentes avanços nas tecnologias de seqüenciamento permitiram a produção de grandes quantidades de dados e redução do custo por base. Bundock *et al*. (2009) mostraram que o uso de NGS é mais rentável que as técnicas anteriormente empregadas para seqüenciamento e identificação de SNPs. NGS foi essencial para o avanço na genômica funcional e vem revolucionando o estudo da genética e as aplicações na prática de melhoramento de plantas (Poland & Rife, 2012; Robasky *et al*., 2014), reduzindo as limitações na geração de informações e facilitando a caracterização de genes e genomas, fornecendo uma visão mais abrangente da diversidade e da função dos genes. Essa técnica vem sendo utilizada para o sequenciamento do genoma inteiro e para projetos de re-sequenciamento, onde os genomas de várias espécies são sequenciados para se descobrir um grande número de polimorfismos SNPs (Elshire *et al*., 2011) a fim de explorar a diversidade genética (Heslot *et al*., 2013; Lu *et al*., 2013; Romay *et al*., 2013), realizar o mapeamento de QTLs (Poland *et al*., 2012; Spindel *et al*., 2013; Liu *et al*., 2014), e o mapeamento associativo

(*Genome-Wide Association Study*, GWAS) (Byrne *et al*., 2013; Crossa *et al*., 2013; Donato *et al*., 2013; Mascher *et al*., 2013; Sonah *et al*., 2013; Uitdewilligen *et al*., 2013; Chen & Lipka *et al*., 2016), além de possibilitar a caracterização de germoplasma.

Novas abordagens para genotipagem baseadas em sequenciamento foram desenvolvidas para aplicar no melhoramento vegetal. Uma dessas abordagens é o GBS, desenvolvido pelo *Institute for Genomic Diversity, Cornell University, USA*. Tal abordagem tem se mostrado um grande aliado para identificação em larga escala de polimorfismos genéticos (Elshire *et al.*, 2011; Lu *et al*., 2013) e pode possibilitar significativo avanço no entendimento da genética da cana-de-açúcar. O protocolo GBS se baseia na redução da complexidade genômica da amostra de DNA total utilizando diferentes combinações de enzimas específicas de restrição (ER) e otimizadas para cada espécie. Após a redução de complexidade via ER e antes do sequenciamento, cada amostra de DNA a ser genotipada recebe adaptadores com sequências indexadoras específicas (*barcodes*) que permitem, mais tarde, rastrear as sequências geradas para cada amostra; desta forma, é possível realizar um multiplex em cada canaleta de sequenciamento, otimizando significativamente os custos (Elshire *et al*., 2011; Poland *et al*., 2012; Glaubitz *et al*., 2014). A filtragem e seleção dos SNPs confiáveis são realizadas seguindo vários critérios. Ao final do procedimento, obtém-se uma quantidade significativa de marcadores SNPs com grau de polimorfismo elevado (Liu *et al*., 2014). Os polimorfismos entre indivíduos podem resultar tanto na presença ou ausência de sequências entre as amostras, o que é derivado da variabilidade na distribuição dos sítios de restrição, ou de SNPs nas sequências em comum entre as amostras. Em cada corrida de sequenciamento podem ser gerados milhares ou dezenas de milhares de marcadores a depender da diversidade nucleotídica da espécie (Lu *et al*., 2013). Além disso, em organismos poliploides, a dosagem alélica dos SNPs e o nível de ploidia de cada loco podem ser estimados através de um modelo gráfico Bayesiano (Figura 3) recentemente desenvolvido para permitir a geração de mapas genéticos de alta densidade (Serang *et al*., 2012; Garcia *et al*., 2013).

As sequências obtidas a partir do GBS são dinâmicas, o que permite que os dados brutos sejam analisados sempre que as técnicas de bioinformática avançarem, possibilitando a descoberta de novos polimorfismos e genes. Outro ponto importante consiste na independência do genoma de referência, fato que contribui enormemente para a sua aplicação em organismos não-modelo, que carecem de informação genômica como no caso da cana-de-açúcar, o qual ainda não possui o sequenciamento do genoma finalizado. A flexibilidade do GBS em relação à espécie, populações e objetivos da pesquisa torna esta uma ferramenta ideal

para o estudo da genética de plantas (Poland *et al*., 2012; Spindel *et al*., 2013; Liu *et al*., 2014). No entanto, algumas limitações devem ser observadas como, por exemplo: i) ocorrência de dados perdidos, imputação e probabilidade de erro na atribuição do genótipo em espécies com alto nível de heterozigosidade; ii) presença regiões repetitivas no DNA dificultam o alinhamento com sequências; iii) densidade de marcadores restrita a regiões específicas do genoma; iv) inconsistência no número de sítios de restrição sequenciados por amostra e no número de *reads* por sítio de restrição; e v) necessidade de elevada profundidade de sequenciamento para organismos poliploides a fim de garantir a estimativa adequada de ploidia e dosagens alélicas dos SNPs identificados (Beissinger et al., 2013; Heffelfinger et al., 2014; Jiang et al., 2016).

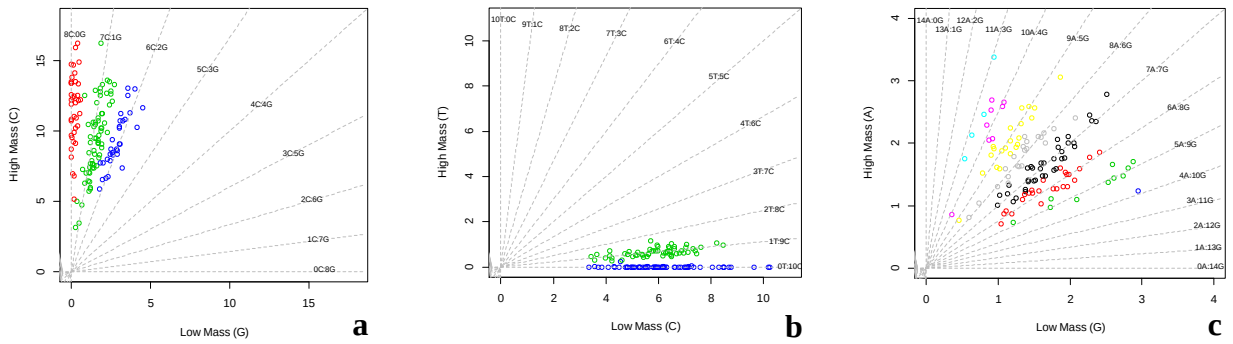


Figura 3 - Estimativa da ploidia e dosagem alélica, através do software SUPERMASSA, para três marcadores SNPs (Fig 3a, Fig 3b e Fig 3c) identificados em base de dados de EST e avaliados em uma população de cana-de-açúcar utilizando a plataforma Sequenom MassARRAY. **Fig 3a:** Marcador SNP estimado com ploidia 8 e dosagem 1:2:1. Para este SNP, ambos genitores da população apresentaram configuração simplex (*CCCCCCCG* X *CCCCCCCG*). **Fig 3b:** Marcador SNP estimado com ploidia 10 e dosagem 1:1. Para este SNP, os genitores da população apresentam configurações nuliplex e simplex (*CCCCCCCCCC* X *CCCCCCCCCT*). **Fig 3c:** Marcador SNP estimado com ploidia 14 e múltiplas doses. Para este SNP, os genitores apresentaram configurações múltiplas (*AAAAAAGGGGGGG* X *AAAAAAAGGGGGG*).

Existe um grande potencial para o uso de SNPs na detecção de associações entre formas alélicas de um gene e fenótipos. Desta forma, com o avanço das tecnologias de sequenciamento e dos modelos genético-estatísticos para lidar apropriadamente com os dados gerados, os marcadores SNPs podem se tornar comumente utilizados também em organismos complexos como a cana-de-açúcar.

Objetivos

Geral

Mapear marcadores microssatélites e TRAP funcionais relacionados a genes que controlam características agroindustriais de importância econômica bem como marcadores SNPs oriundos de genotipagem por sequenciamento (*Genotyping by Sequencing*, GBS) numa progênie derivada do cruzamento entre duas cultivares comerciais de cana-de-açúcar.

Específicos

- Avaliar o polimorfismo de 400 marcadores microssatélites nos parentais da população de mapeamento derivada do cruzamento entre SP80-3280 e RB835486;

- Detectar marcadores SNPs a partir de genotipagem por sequenciamento;

- Construir um mapa genético molecular utilizando uma progênie de interesse comercial;

- Analisar características economicamente importantes e mapear QTLs envolvidos no controle destas características;

- Comparar o mapa gerado para a progênie do cruzamento entre SP80-3280 e RB835486 com mapas de outras populações já publicados, com o intuito de validar os marcadores funcionais associados com as características de interesse, permitindo a aplicação destes marcadores na seleção assistida.

CAPÍTULO 1

Mixed Modeling of Yield Components and Brown Rust Resistance in Sugarcane Families

Thiago W. A. Balsalobre, Melina C. Mancini, Guilherme da S. Pereira, Carina de O. Anoni, Fernanda Z. Barreto, Hermann P. Hoffmann, Anete P. de Souza, Antonio A. F. Garcia, and Monalisa S. Carneiro*

T.W.A. Balsalobre, F.Z. Barreto, H.P. Hoffmann, and M.S. Carneiro, Center for Agricultural Sciences, Dep. of Biotechnology and Plant and Animal Production, Federal Univ. of São Carlos, Highway Anhanguera Km 174, Araras, CEP 13600-970, SP, Brazil; M.C. Mancini and A.P. Souza, Dep. of Plant Biology, Biology Institute, Univ. of Campinas, 255 Monteiro Lobato Ave, Campinas, CEP 13083-862, SP, Brazil, and Center for Molecular Biology and Genetic Engineering, 400 Candido Rondon Ave, Campinas, CEP 13083-875, SP, Brazil; and G.S. Pereira, C.O. Anoni, and A.A.F. Garcia, Dep. of Genetics, Luiz de Queiroz College of Agriculture, Univ. of São Paulo, 11 Pádua Dias Ave, Piracicaba, CEP 13418-900, SP, Brazil. T.W.A. Balsalobre and M.C. Mancini contributed equally to this work. *Corresponding author (monalisa@cca.ufscar.br).

Published September 8, 2016

BIOMETRY, MODELING & STATISTICS

Mixed Modeling of Yield Components and Brown Rust Resistance in Sugarcane Families

Thiago W. A. Balsalobre, Melina C. Mancini, Guilherme da S. Pereira, Carina O. Anoni, Fernanda Z. Barreto, Hermann P. Hoffmann, Anete P. de Souza, Antonio A. F. Garcia, and Monalisa S. Carneiro*

ABSTRACT

Sugarcane (*Saccharum* spp.) is a complex autopolyploid with high potential for biomass production that can be converted into sugar and ethanol. Genetic improvement is extremely important to generate more productive and resistant cultivars. Populations of improved sugarcane are generally evaluated for several traits simultaneously and in multi-environment trials. In this study, we evaluated two full-sib families of sugarcane (SR1 and SR2) at two locations and 3 yr for stalk diameter, stalk height, stalk number, stalk weight, soluble solid content (Brix), sucrose content of cane, sucrose content of juice, fiber, cane yield, sucrose yield, and resistance to brown rust (*Puccinia melanocephala*). Using a mixed model approach, we included appropriate variance–covariance (VCOV) structures for modeling heterogeneity and correlation of genetic effects and non-genetic residual effects. The genotypic correlations between traits were calculated across the adjusted means as the standard Pearson product-moment coefficient. Through the VCOV structures estimated for each trait, in general, the heritabilities ranged from 0.78 to 0.94. Additionally, we detected 17 and 12 significant genotypic correlations between the evaluated traits for SR1 and SR2, respectively. The analysis of the severity data for brown rust revealed that 66 and 32% of the full-sib genotypes in SR1 and SR2, respectively, had at least 90% probability of being resistant.

Core Ideas

- A linear mixed model is efficient in production data analysis of sugarcane.
- In general, the broad-sense heritability of the traits were high, ranging from 0.78 to 0.94.
- A generalized linear mixed model can be applied in brown rust analysis of sugarcane.
- Multi-environment trials were applied to the genetic improvement of sugarcane.

Published in Agron. J. 108:1824–1837 (2016)
doi:10.2134/agronj2015.0430
Received 4 Sept. 2015
Accepted 13 Feb. 2016
Supplemental material available online
Available freely online through the author-supported open access option

Sugarcane is a complex autopolyploid plant and one of the most highly consumed crops in the world (FAO, 2014). Sugarcane produces high yields and has been demonstrated to efficiently use resources (i.e., land, water, N, and energy) (de Vries et al., 2010; Eksteen et al., 2014; Gerbens-Leenes et al., 2009; Waclawovsky et al., 2010). A sustainable energy future depends on the increased use of renewable energy. Bioethanol produced from sugarcane through both first- and second-generation technologies is a good example of a renewable energy source that can contribute to reducing the environmental effects of fossil fuels (Goldemberg, 2007). First-generation production is an alternative and economically viable technique that is widely used in Brazil, whereas second-generation production is not as well optimized as first-generation production and is thus less feasible. However, the substantial potential of the generation of bioethanol from lignocellulosic wastes has encouraged studies that seek to improve the process and make it an integral component of the units that already produce first-generation bioethanol (Macrelli et al., 2014; Naik et al., 2010; Saini et al., 2015).

Sugarcane breeding programs have focused on releasing new cultivars with agronomic traits that suit the demand of the sugarcane industry. However, the genetic complexity of sugarcane has hindered the genetic improvement of this crop. Commercial sugarcane cultivars are the result of interspecific crosses between both domesticated *Saccharum officinarum* L. ($2n = 80$) and wild *S. spontaneum* ($2n = 40–120$) species, followed by several

T.W.A. Balsalobre, F.Z. Barreto, H.P. Hoffmann, and M.S. Carneiro, Center for Agricultural Sciences, Dep. of Biotechnology and Plant and Animal Production, Federal Univ. of São Carlos, Highway Anhanguera Km 174, Araras, CEP 13600-970, SP, Brazil; M.C. Mancini and A.P. Souza, Dep. of Plant Biology, Biology Institute, Univ. of Campinas, 255 Monteiro Lobato Ave, Campinas, CEP 13083-862, SP, Brazil, and Center for Molecular Biology and Genetic Engineering, 400 Candido Rondon Ave, Campinas, CEP 13083-875, SP, Brazil; and G.S. Pereira, C.O. Anoni, and A.A.F. Garcia, Dep. of Genetics, Luiz de Queiroz College of Agriculture, Univ. of São Paulo, 11 Pádua Dias Ave, Piracicaba, CEP 13418-900, SP, Brazil. T.W.A. Balsalobre and M.C. Mancini contributed equally to this work. *Corresponding author (monalisa@cca.ufscar.br).

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; BLUP, best linear unbiased prediction; FIB, fiber; GEI, genotype × environment interaction; GLMM, generalized linear mixed model; LMM, linear mixed model; METs, multi-environment trials; POL%C, sucrose content of cane in percentage; POL%J, sucrose content of juice in percentage; REML, restricted maximum likelihood; SD, stalk diameter; SH, stalk height; SN, stalk number; SR1, SP80-3280 × RB835486 full-sib family of sugarcane 1; SR2, SP81-3250 × RB925345 full-sib family of sugarcane 2; SW, stalk weight; TCH, tonnes of cane per hectare; TPH, tonnes of sucrose per hectare; VCOV, variance–covariance.

backcrosses with *S. officinarum* (Ha et al., 1999; Irvine, 1999). The polyploid genome complexity of commercial cultivars can be attributed to the following factors: the chromosome number, ranging from 100 to 130 (D'Hont et al., 1998; Irvine, 1999); the genome size of approximately 10 Gb (D'Hont, 2005; D'Hont and Glaszmann, 2001; Piperidis et al., 2010); and the aneuploidy condition, with a variable number of chromosomes in each hom(e)ology group (Grivet and Arruda, 2002).

The primary purpose of a genetic breeding program is to improve yields (Cox et al., 1994), which is possible due to the accumulation of knowledge of plant breeding. Worldwide data indicate that sugarcane yield has increased by 41% during the last 50 yr (Gouy et al., 2013, 2015). Recently, small sugar production increases of approximately 1 to 1.5% per year have been obtained. The average yield of sugarcane in Brazil, the world's largest sugarcane producer, is approximately 74 t ha^{-1}; however, the theoretical production potential is approximately 400 t ha^{-1} (Dal-Bianco et al., 2012; Matsuoka et al., 2014; Waclawovsky et al., 2010). Waclawovsky et al. (2010) showed that, in Brazil, the commercial maximum yield (large land areas) was 260 t ha^{-1} and an experimental maximum (individual trials on smaller land areas) was 299 t ha^{-1}. These high yields were obtained under irrigation in an area with low precipitation and low cloudiness, hence higher solar radiation than is observed in most sugarcane-producing areas of Brazil. Thus, to achieve high yields, it is necessary to consider the use of agricultural practices that involve additional costs, such as irrigation and fertilization. The genetic components, the environment, and the relationship between the traits of interest are essential to developing breeding strategies. The genetic component of the phenotypic variance in crop traits is most commonly studied, as reflected in the high rate of scientific and technological progress in plant breeding (Edwards et al., 2013; Sadras et al., 2013). In sugarcane breeding programs, many genetic clones are commonly evaluated during several harvests in multi-environment trials (METs). Genotype × environment interaction (GEI) is broadly considered to be the variation in the relative performance of genotypes across environments (Ramburan et al., 2012) and is an important feature when selecting superior cultivars (Jackson et al., 1991; Ramburan, 2014). Furthermore, experiments commonly involve the simultaneous evaluation of several traits because superior cultivars should concentrate favorable alleles for yield, resistance to diseases (e.g., brown rust and smut), pests and abiotic stresses, and agronomic traits, among other factors (Welham et al., 2010).

Brown rust, a disease caused by *Puccinia melanocephala* H. & P. Sydow, affects sugarcane and is present in many production areas worldwide (Asnaghi et al., 2004; Hoy and Hollier, 2009; Ryan and Egan, 1989). Negative impacts of brown rust on sugarcane yields have been reported (McFarlane et al., 2006; Raid and Comstock, 2000). Field losses >50% are associated with brown rust, depending on the cultivar susceptibility and growing conditions (Hoy and Hollier, 2009; Purdy et al., 1983). Therefore, the release of cultivars that are resistant to brown rust is very important and is the most efficient form of disease control. The genetic inheritance of brown rust in sugarcane has been broadly studied, and some researchers have claimed that this resistance trait is controlled by one or a few genes (Asnaghi et al., 2004; Costet et al., 2012; Daugrois et al., 1996; Garsmeur et al., 2011; Glynn et al., 2013; Hogarth et al., 1993; Le Cunff et al., 2008; Parco et al., 2014; Raboin et al., 2006; Racedo et al., 2013; Ramdoyal et al., 2000; Sordi et al., 1988). In contrast to that shown for brown rust, almost every trait that is related to sugarcane production exhibits quantitative variations. For example, sugar yield components depend on a combination of stalk diameter, stalk height, stalk number, stalk weight, and BRIX (Hogarth, 1971). These traits have a complex relationship, which complicates the selection of superior cultivars.

Because of the combination of factors for a sugarcane cultivar ideotype, the efficient estimation of genetic parameters is dependent on the choice of experimental design and statistical models that are appropriate for the response pattern of the evaluated variables (Sadras et al., 2013). In traditional analysis of variance models, all of the effects are considered fixed (except for the residual error), limiting the potential of the analysis. For sugarcane, the data from METs are modeled by assuming variance homogeneity and absence of genetic correlation between the harvest and location for estimating breeding values (Balzarini, 2002). In contrast, linear mixed models (LMMs) (Henderson, 1984) have advantages over fixed linear models for analyzing METs. Specifically, LMMs have the ability to consider variables as random rather than fixed and to use different variance–covariance (VCOV) structures for random effects to investigate the presence of heteroscedasticity and correlations. This approach allows the analysis of unbalanced data (Pastina et al., 2012; Smith et al., 2005) in addition to using more realistic models for residual variation (incomplete blocks and spatial correlation) and assuming sets of effects (e.g., genotypes) as random (Piepho et al., 2008; Smith et al., 2005). The estimation of variance component parameters is obtained preferably by restricted maximum likelihood (REML), and genotype effects may be obtained either by best linear unbiased estimation or best linear unbiased prediction (BLUP), depending on whether genotypes are considered fixed or random factors, respectively (Piepho et al., 2008; Smith et al., 2005). One major property of BLUP is shrinkage toward the mean, which anticipates regression of progeny to the mean and increases the accuracy of prediction of breeding and genotypic values (Piepho et al., 2008). In addition, BLUP maximizes the correlation of true genotypic values and predicted genotypic values, which is the primary aim of breeders (Searle et al., 1992).

The mixed model approach is more realistic, with a higher predictive ability based on modeling the VCOV matrices. This approach also means a great change in the analysis of breeding experiments because genotype observations may be grouped by levels of grouping factors generated from the experimental design, such as the harvest year and location (Pastina et al., 2012). The application of a mixed model approach is becoming increasingly popular in plant breeding, particularly in research involving the prediction of breeding values combined with genomic data (Beaulieu et al., 2014; Bevan and Uauy, 2013; Burgueño et al., 2012; Crossa et al., 2013; Muir, 2007; Wolc et al., 2011; Zhang et al., 2010). For sugarcane, the use of linear mixed models to map quantitative trait loci (Pastina et al., 2012) and genomic selection (Gouy et al., 2013) represents progress in crop improvement. However, the basis of genetic inheritance of all traits of economic interest and the genetic expression of related genes in cultivars that are used as parents by breeding programs should be better understood so that we

can efficiently achieve higher gains with the selection process and apply genomic selection in sugarcane in the future.

The objectives of this study were to: (i) evaluate the morphological and technological traits in full-sib families of sugarcane that were established in two different locations during three harvest years using LMMs, (ii) estimate the heritability and genotypic correlation coefficients using VCOV matrices, and (iii) assess the severity data of brown rust disease using a generalized linear mixed model (GLMM).

MATERIALS AND METHODS

Plant Material

The populations were developed by the Genetic Breeding Program of Sugarcane from the Universidade Federal de São Carlos (UFSCar), which is an integral component of the Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RIDESA). Bi-parental crosses between Brazilian commercial cultivars produced two full-sib families from which the phenotypic data were collected. The first family, named SR1, consisted of 153 full-sib genotypes that were derived from a cross between SP80-3280 (female parent) and RB835486 (male parent). The SP80-3280 parent (SP71-1088 × H57-5028) was sequenced by the Sugarcane Expressed Sequence Tags (SUCEST) Project and has higher productivity, sucrose content, fiber content, and resistance to smut (*Sporisorium scitamineum* Sydow) and brown rust; RB835486 (L60-14 × ?) has a higher sucrose content and is susceptible to smut and brown rust. The second family, named SR2, consisted of 240 full-sib genotypes that were derived from a cross between SP81-3250 (female parent) and RB925345 (male parent). The SP81-3250 parent (CP70-1547 × SP71-1279) is resistant to brown rust, while RB925345 (H59-1966 × ?) is susceptible to brown rust; both parents have high productivity, sucrose content, and fiber content.

Phenotypic Data

Two independent experiments, one for each family, were planted in 2010 at two locations (Araras and Ipaussu) in the state of São Paulo, Brazil. The Araras site was located at 22°21′25″ S, 47°23′3″ W, at an altitude of 611 m; the soil of the site was a Typic Eutroferric Red Latosol. The Ipaussu site was located at 23°8′44″ S, 49°23′23″ W, at an altitude of 477 m; the soil of the site was a Dark Red Latosol. Historically, Ipaussu is a location with a high natural incidence of brown rust.

At each location, the experimental design consisted of an augmented randomized incomplete block design, which was fully replicated three times. Each incomplete block included 30 genotypes: 27 full-sib genotypes plus three checks (SP80-3280, RB835486, and RB867515 for the SR1 experiment and SP81-3250, RB925345, and RB867515 for the SR2 experiment). The positions of the 30 genotypes in each incomplete block were fully randomized within family. Trial plots consisted of three and two rows in Ipaussu and Araras, respectively. The rows were 3 m long and spaced 1.5 m apart for both locations.

Sugarcane families were evaluated for 10 yield components: soluble solid content (BRIX, in °Brix), sucrose content of the cane (POL%C, in %), sucrose content of the juice (POL%J, in %), fiber (FIB, in %), stalk diameter (SD, in mm), stalk height (SH, in cm), stalk number (SN), stalk weight (SW, in kg), cane yield (TCH, in t ha^{-1}), and sucrose yield (TPH, in t ha^{-1}). Considering each experiment as multi-harvest-location, trials were harvested when the plants were approximately 12 mo of age during 2011, 2012, and 2013 at both locations. A 10-stalk sample was taken for analysis of the BRIX, POL%C, POL%J, and FIB, and two replicates at each location were evaluated. For analysis of the SD, SH, SN, SW, and TCH, all three replicates at each location were evaluated. A cluster of 10 stalks per plot was weighed and used to measure SH and SD. The weights of the two clusters of 10 stalks from each plot were added to the total weight of the plot (SW) to estimate the TCH, which was calculated as the product of the stalk weight of a linear meter (6667 linear meters compose 1 ha with a spacing of 1.5 m). The number of stalks was estimated by directly counting the tillers in the field. The values of TCH and POL%C were used to estimate TPH from the product of TCH and POL%C divided by 100. The yield component data were evaluated at a plant age of 12 mo according to the methodology described by the State of Sao Paulo Sugarcane, Sugar and Alcohol Growers Council (2006). During the experiments, Family SR1 suffered unforeseen events (the experimental field was attacked by capybaras [*Hydrochoerus hydrochaeris*]) that rendered the collection of yield component data of the second harvest at Araras unviable. Likewise, for both families (SR1 and SR2), the collection of BRIX, POL%C, POL%J, and FIB data for the second harvest at Ipaussu was not possible because there was an accidental fire in the experimental field that prevented the collection of samples for these analyses.

The resistance to brown rust was evaluated in the field through natural infestation. Evaluations of the incidence of brown rust were performed on both full-sib genotypes and checks on 6-mo-old plants in February during 2011, 2012, and 2013. This period is considered the most favorable epidemiological season for the occurrence of brown rust, considering both temperature and humidity conditions favoring the incidence of infection as the plants age. Brown rust resistance was scored in 3+ on each plot on a 1 (most resistant) to 9 (most susceptible) diagrammatic scale according to Tai et al. (1981) and Amorim et al. (1987). This scale is based on a visual assessment of the disease symptoms. A score of 1 indicates the absence of sporulating pustules (uredospores) and resistant plants. A score of 2 indicates very rare sporulating pustules. For grades from 2 to 9, the density of sporulating pustules increases and indicates susceptible plants. Five plants per plot were evaluated.

Analysis of Yield Components and Brown Rust Resistance Data

A multi-harvest-location mixed model produced the joint-adjusted means to obtain genotypic correlations among traits. The analyses were conducted for each trait for both populations using GenStat 16 (Payne et al., 2009) based on the REML and the following linear model:

$$y_{ijkmn} = \mu + l_n + h_m + r_{k(nm)} + b_{j(kmn)} + t_{imn} + \varepsilon_{ijkmn} \quad [1]$$

where y_{ijkmn} is the phenotype of the ith genotype in the kth replicate and the jth incomplete block at the nth location and mth harvest; μ is the overall mean; l_n is the fixed effect of the nth location ($n = 1, N = 2$); h_m is the fixed effect of the mth harvest ($m = 1, ..., M$; $M = 2$ or 3 depending on the location);

$r_{k(nm)}$ is the fixed effect of the kth replicate ($k = 1, ..., K$; $K = 2$ or 3 depending on the trait) at the nth location and mth harvest; $b_{j(knm)}$ is the random effect of the jth block ($j = 1, ..., J$; $J = 9$) in the kth replicate at the nth location and mth harvest; t_{inm} is the random effect of the ith genotype ($i = 1, ..., I$; $I = 156$ for SR1 and $I = 243$ for SR2) at the nth location and mth harvest; and ε_{ijknm} is the random residual error. The genotypes (t_{inm}) were separated into two groups, in which g_{inm} was a random genetic effect of the ith full-sib genotype ($i = 1, ..., I_g$; $I_g = 153$ for SR1 or 240 for SR2) at the nth location and mth harvest, and c_{inm} was the fixed effect of the ith check ($i = I_g + 1, ..., I_g + I_c$; $I_c = 3$) at the nth location and mth harvest. For the genotypes, the vector $\mathbf{g} = (g_{111}, ..., g_{IMN})'$ had a multivariate normal distribution with zero mean vector and genetic VCOV matrix $\mathbf{G} = \mathbf{G}_P \otimes \mathbf{I}_{I_g}$, i.e., $\mathbf{g} \sim N(0,\mathbf{G})$, where P is the number of location–harvest combinations and $\otimes$ represents the Kronecker direct product of both the genetic $\mathbf{G}_P$ and identity $\mathbf{I}_{I_g}$ matrices with the respective dimensions of $P \times P$ and $I_g \times I_g$. Several structures for the $\mathbf{G}_P$ matrix (Table 1) were examined and compared via Akaike (AIC; Akaike, 1974) and Bayesian (BIC; Schwarz, 1978) information criteria (Pastina et al., 2012). Abbreviations of each structure presented in Table 1 will be used hereafter in the text. Models 1 to 6 used a location–harvest factorial combination for different environments (E), i.e., $\mathbf{G}_P = \mathbf{G}^{E}_{P\times P}$, whereas Models 7 to 12 used VCOV matrix direct products for location (L) and harvest (H), i.e., $\mathbf{G}_P = \mathbf{G}^{L}_{N\times N} \otimes \mathbf{G}^{H}_{M\times M}$. For residuals, $\varepsilon \sim N(0,\mathbf{R})$, where $\varepsilon = (\varepsilon_{11111}, ..., \varepsilon_{IJKMN})'$, and $\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{I\,J}$ is the residual VCOV matrix, whereas matrices $\mathbf{R}_P = \mathbf{R}^{E}_{P\times P}$ (factorial combination) or $\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$ (direct product) and $\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$ were examined and compared via AIC and BIC for several structures of locations, harvest, and replicates after the selection of $\mathbf{G}_P$. The selected $\mathbf{R}$ matrices were included in the final phenotypic models by considering the existence of non-genetic residual correlations and the heterogeneity of non-genetic residual variances in all harvests and locations. Based on these models, the adjusted means for individual traits in each family and the genetic parameters could be obtained. For each trait, the fixed effects of the interactions between the location, harvest, and checks were tested using the Wald statistics test and were retained in the model if statistically significant ($P < 0.05$). The genotypic correlations between the traits were calculated across the adjusted means as the standard Pearson product-moment coefficient and were tested by assuming a significant global level of $\alpha^* = 0.05$ in R software (http://www.cran.r-project.org) using the package *psych* (Revelle, 2014), which was also used to draw scatterplots between pairs of traits. The broad-sense heritabilities on an individual-plant basis ($\hat{H}_{plants}$) were computed based on variance component estimates assuming an identity structure for the $\mathbf{G}_P$ matrix (Model 1) using the ratio $\hat{\sigma}^2_G / \hat{\sigma}^2_P$, where $\hat{\sigma}^2_G$ is the among-genotype variance component and $\hat{\sigma}^2_P$ is the total phenotypic variance for each trait. The ratio $\hat{\sigma}^2_G / \hat{\sigma}^2_P$ was computed to provide approximate measurements of the broad heritabilities on a genotype-mean basis ($\hat{H}_{means}$), where $\hat{\sigma}^2_P$ is the phenotypic variance among the genotype means for each trait obtained using the harmonic mean of the number of environments as the numerator of the GEI variance estimates and the harmonic mean of the number of full-sib genotypes sampled in each experiment as the numerator of the residual error variance estimates (Holland et al., 2003).

The brown rust severity data did not follow a Gaussian distribution. As a first approach to the problem, a GLMM was used for this analysis (Gianola and Foulley, 1983; Gouy et al., 2013; Thompson, 1979). The average severity of each genotype was transformed into a binary scale, where 0 represents disease resistance (1 on the diagrammatic scale) and 1 represents susceptibility to disease (2–9 on the diagrammatic scale). The transformation of the continuous variable data related to the severity of brown rust in two classes, resistance and susceptibility, was also reported by Asnaghi et al. (2004), Raboin et al. (2006), and Costet et al. (2012).

Table 1. Description and number of parameters (n_{PAR}) of the examined models for the genetic variance-covariance matrix $\mathbf{G}_P$ (Models 1–6 used the factorial combination of locations and harvests as different environments [E]; Models 7–12 used the direct product of covariance matrices for locations [L] and harvests [H]; $P = NM$, where N is the number of locations, and M is the number of harvests).

Model	Parameter	n_{PAR}	Description
		$\mathbf{G}_P = \mathbf{G}^{E}_{P\times P}$	
1	ID	1	identity (or homogeneous genetic variances)
2	UNIF	2	uniform
3	DIAG	P	diagonal (or heterogeneous genetic variances)
4	CS_{Het}	$P + 1$	compound symmetry with heterogeneous genetic variation
5	FA1	$2P$	first-order factor analytic
6	UNST	$P(P + 1)/2$	unstructured
		$\mathbf{G}_P = \mathbf{G}^{L}_{N\times N} \otimes \mathbf{G}^{H}_{M\times M}$	
7	UNST ⊗ ID	$N(N + 1)/2 + 1$	unstructured and identity for locations and harvest, respectively
8	UNST ⊗ UNIF	$N(N + 1)/2 + 2$	unstructured and uniform for locations and harvest, respectively
9	UNST ⊗ DIAG	$N(N + 1)/2 + M$	unstructured and diagonal for locations and harvest, respectively
10	UNST ⊗ AR1	$[N(N + 1) + 2(M + 1)]/2 - 1$	unstructured and first-order autoregressive for locations and harvest, respectively
11	UNST ⊗ CS_{Het}	$N(N + 1)/2 + M + 1$	unstructured and compound symmetry for locations and harvest, respectively
12	UNST ⊗ UNST	$[N(N + 1) + M(M + 1)]/2 - 1$	unstructured for both locations and harvest

A GLMM with a binomial error distribution and a logit link function was used to model the underlying susceptibility to the disease (observed binary phenotype) (Bolker et al., 2009; De Silva et al., 2014). Location, harvest, and replicate terms were treated as fixed effects, and their significances were assessed by Wald test statistics ($P < 0.05$). A random genotype effect was incorporated into the model and was assumed to be normally distributed, with a zero mean and variance component $\hat{\sigma}^2_G$. The REML was used to estimate the model parameters and variance components using the method of Trust (2014), as implemented in GenStat 16 software (Payne et al., 2009). The approximate broad-sense heritability on a plant-mean basis ($\hat{H}^2_{plants}$) was computed based on $\hat{\sigma}^2_G / \{\hat{\sigma}^2_G + \hat{\sigma}^2\}$, where $\hat{\sigma}^2_G$ is the genetic variance component estimate and $\hat{\sigma}^2$ is the dispersion parameter estimate.

RESULTS

Model Selection for Multi-Harvest-Location Yield Components

Several structures for the $\mathbf{G}_P$ matrix examined and compared via AIC and BIC are summarized in Supplemental Table S1. Four models for the $\mathbf{G}_P$ matrix that consider the heterogeneity of variance were selected according to the data of the evaluated traits (FA1, UNST, UNST ⊗ AR1, and UNST ⊗ UNST) and are provided in Table 2 for each family, SR1 and SR2. The matrix selection for random effects showed that for SR1, the selected models for the traits of SD, SW, BRIX, POL%J, FIB, TCH, and TPH considered the VCOV structure $\mathbf{G}_P = \mathbf{G}^{L}_{N\times N} \otimes \mathbf{G}^{H}_{M\times M}$. The model that was selected for the traits of SN, SH, and POL%C considered the VCOV structure $\mathbf{G}_P = \mathbf{G}^{E}_{P\times P}$. For SR2, the selected models for most traits (SD, SH, SN, BRIX, POL%C, POL%J, FIB, TCH, and TPH) considered a VCOV structure of $\mathbf{G}_P = \mathbf{G}^{E}_{P\times P}$. The selected model considered the VCOV structure of $\mathbf{G}_P = \mathbf{G}^{L}_{N\times N} \otimes \mathbf{G}^{H}_{M\times M}$ only for the SW trait. For each trait, the selection of the mean structure (fixed part of the model) using Wald statistics for SR1 indicated that the interaction effects between harvest and checks were not significant for SH, SN, BRIX, POL%C, POL%J, FIB, TCH, and TPH and that the interaction effects between location and checks were not significant for SD, POL%C, and FIB. In contrast, for SR2, the Wald statistics showed that the interactions between harvest and checks were not significant for SD and FIB and that the interactions between location and checks were not significant for BRIX, POL%C, POL%J, and FIB. On the other hand, the interactions between the location and harvest were not significant for POL%C. The nonsignificant effects were removed from the model, and then the adjusted means were obtained.

For non-genetic residual effects, structures for the **R** matrix examined and compared via AIC and BIC are summarized in Supplemental Table S2 for SR1 and Supplemental Table S3 for SR2. The models that were selected according to the data of the evaluated traits can be viewed in Table 3 for each family, SR1 and SR2. The data of the evaluated traits that fit the factorial combination among locations and harvests in SR1 (SH, SN, POL%C) and SR2 (SD, SH, SN, BRIX, POL%C, POL%J, FIB, TCH, and TPH) showed a pattern of homoscedasticity to nongenetic residual effects between environments, except for POL%C in SR1 and SD, SN, FIB, and TCH in SR2. Furthermore, for these traits that fit the factorial combination among locations and harvests, the covariance between environments was null in SR2 except for the traits SH, SN, TCH, and TPH, whereas in SR1, the covariance was the same between pairs of environments for SH and SN and different for POL%C. For both families, SH and SW showed the same pattern of nongenetic residual effects: SH presented homoscedasticity between environments with the same covariance between pairs of environments, and SW showed heteroscedasticity with null covariance between locations and heteroscedasticity with different covariance between harvests. Every other evaluated trait showed different patterns of behavior for nongenetic residual effects between families.

Table 2. Selected models for the $\mathbf{G}_P$ matrix and number of estimated parameters (n_{PAR}) considering each trait separately. The Akaike (AIC) and Bayesian (BIC) information criteria were used to compare the structures of the variance–covariance matrix. The models for the G_P matrix were selected according to the lowest value of the AIC for the stalk diameter (SD) in mm, stalk height (SH) in m, stalk number (SN) by direct counting, stalk weight (SW) in kg, BRIX as °Brix, sucrose content of cane (POL%C) in percentage, sucrose content of juice (POL%J) in percentage, fiber (FIB) as a percentage, cane yield (TCH) in t ha^{-1}, and sucrose yield (TPH) in t ha^{-1} for the two families of sugarcane (SR1 and SR2) at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012, and 2013).

Trait	Selected model for $\mathbf{G}_P$ matrix†	n_{PAR}	AIC	BIC
	SP80–3280 × RB835486 (SR1)			
SD, mm	10. UNST ⊗ AR1	5	18,930.4	18,968.1
SH, m	5. FA1	10	1274.4	1349.8
SN	5. FA1	10	37,808.5	37,884.0
SW, kg	10. UNST ⊗ AR1	5	42,491.7	42,529.4
BRIX °Brix	12. UNST ⊗ UNST	6	6439.5	6479.2
POL%C	6. UNST	10	6407.1	6475.0
POL%J	12. UNST ⊗ UNST	6	7109.1	7148.8
FIB, %	12. UNST ⊗ UNST	6	6050.5	6090.1
TCH, t ha^{-1}	10. UNST ⊗ AR1	5	36,511.5	36,549.2
TPH, t ha^{-1}	12. UNST ⊗ UNST	6	11,646.2	11,685.8
	SP81–3250 × RB925345 (SR2)			
SD, mm	6. UNST	10	20,096.9	20,245.3
SH, m	6. UNST	10	728.2	876.7
SN	5. FA1	10	45,429.0	45,519.6
SW, kg	12. UNST ⊗ UNST	6	47,036.9	47,101.5
BRIX, °Brix	6. UNST	10	7767.2	7867.1
POL%C	6. UNST	10	7653.2	7753.0
POL%J	6. UNST	10	8637.5	8737.4
FIB, %	6. UNST	10	6630.1	6730.0
TCH, t ha^{-1}	6. UNST	10	45,244.1	45,392.7
TPH, t ha^{-1}	6. UNST	10	15,411.7	15,546.7

† Models selected for the $\mathbf{G}_P$ matrix as described in Table 1.

Heritability and Components of Variance in Yield Components

The results regarding the ranges, averages of families, averages of parents, estimates of the components of variance, coefficient of variation, and the broad-sense heritability on an individual-plant and genotype-mean basis of the 10 traits evaluated for the two families are summarized in Table 4. In general, the $\hat{H}^2_{means}$ of the traits were high (>0.80) for both families. The $\hat{H}^2_{means}$ ranged from 0.78 (SH) to 0.92 (SD) in SR1 and from 0.79 (POL%C) to

Table 3. Selected models for the R matrix and number of estimated parameters (n_{PAR}) considering each trait separately. Akaike (AIC) and Bayesian (BIC) information criteria were used to compare the structures of the variance–covariance matrix. The models for the R matrix were selected according to the lowest value of the BIC for the stalk diameter (SD) in mm, stalk height (SH) in m, stalk number (SN) by direct counting, stalk weight (SW) in kg, BRIX as °Brix, sucrose content of cane (POL%C) in percentage, sucrose content of juice (POL%J) in percentage, fiber (FIB) as a percentage, cane yield (TCH) in t ha^{-1}, and sucrose yield (TPH) in t ha^{-1} for the two families of sugarcane (SR1 and SR2) at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012, and 2013).

Trait	Selected model for R matrix	n_{PAR}			AIC			BIC		
		SP80–3280 × RB835486 (SR1)								
	$\mathbf{R} = \mathbf{R}^{E}_{P\text{-}P} \otimes \mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$		$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$		$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$	
SH, m	$\mathbf{R}$ = UNIF ⊗ ID	2	1		1266.1	1266.1		1347.8	1347.8	
SN	$\mathbf{R}$ = UNIF ⊗ ID	2	1		37,534.3	37,534.3		37,616.0	37,616.0	
POL%C	$\mathbf{R}$ = UNST ⊗ ID	10	1		6304.7	6304.7		6423.6	6423.6	
	$\mathbf{R} = \mathbf{R}^{L}_{N\text{-}N} \otimes \mathbf{R}^{H}_{M\text{-}M} \otimes \mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$
SD, mm	$\mathbf{R}$ = DIAG ⊗ CS_{Het} ⊗ DIAG	2	4	3	18,906.2	18,716.2	18,653.8	18,950.1	18,779.1	18,729.2
SW, kg	$\mathbf{R}$ = DIAG ⊗ UNST ⊗ CS_{Het}	2	6	4	41,985.3	41,608.4	41,371.4	42,029.3	41,683.9	41,645.7
BRIX, °Brix	$\mathbf{R}$ = ID ⊗ DIAG ⊗ ID	1	3	1	6439.5	6323.4	6428.8	6479.2	6374.4	6474.1
POL%J	$\mathbf{R}$ = ID ⊗ UNIF ⊗ ID	1	2	1	7109.1	7002.5	7002.5	7148.8	7047.9	7047.9
FIB, %	$\mathbf{R}$ = ID ⊗ UNST ⊗ DIAG	1	3	2	6050.5	5923.5	5912.1	6090.1	5974.5	5968.7
TCH, t ha^{-1}	$\mathbf{R}$ = UNST ⊗ UNIF ⊗ DIAG	3	2	3	36,498.7	35,980.8	35,961.1	36,549.0	36,037.4	36,030.3
TPH, t ha^{-1}	$\mathbf{R}$ = ID ⊗ UNIF ⊗ DIAG	1	2	2	11,646.2	11,496.5	11,490.4	11,685.8	11,541.8	11,541.3
		SP81–3250 × RB925345 (SR2)								
	$\mathbf{R} = \mathbf{R}^{E}_{P\text{-}P} \otimes \mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$		$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$		$\mathbf{R}^{E}_{P\text{-}P}$	$\mathbf{R}^{R}_{K\text{-}K}$	
SD, mm	$\mathbf{R}$ = DIAG ⊗ UNST	6	6		19,508.9	19,335.7		19,689.7	19,548.8	
SH, m	$\mathbf{R}$ = UNIF ⊗ ID	2	1		702.6	702.6		857.5	857.5	
SN	$\mathbf{R}$ = UNST ⊗ CS_{Het}	21	4		43,966.4	43,943.4		44,186.2	44,186.2	
BRIX, °Brix	$\mathbf{R}$ = ID ⊗ ID	1	1		7767.2	7767.2		7867.1	7867.1	
POL%C	$\mathbf{R}$ = ID ⊗ ID	1	1		7653.2	7653.2		7753.0	7753.0	
POL%J	$\mathbf{R}$ = ID ⊗ ID	1	1		8637.5	8637.5		8737.4	8737.4	
FIB, %	$\mathbf{R}$ = DIAG ⊗ ID	5	1		6489.8	6489.8		6613.1	6613.1	
TCH, t ha^{-1}	$\mathbf{R}$ = UNST ⊗ UNIF	21	2		44,621.9	44,605.1		44,899.7	44,889.3	
TPH, t ha^{-1}	$\mathbf{R}$ = UNIF ⊗ ID	2	1		15,319.9	15,319.9		15,460.7	15,460.7	
	$\mathbf{R} = \mathbf{R}^{L}_{N\text{-}N} \otimes \mathbf{R}^{H}_{M\text{-}M} \otimes \mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$	$\mathbf{R}^{L}_{N\text{-}N}$	$\mathbf{R}^{H}_{M\text{-}M}$	$\mathbf{R}^{R}_{K\text{-}K}$
SW, kg	$\mathbf{R}$ = DIAG ⊗ UNST ⊗ UNST	2	3	6	46,497.2	45,954.0	46,065.3	46,568.3	46,057.4	45,929.5

0.94 (SD) in SR2. Estimates for $\hat{h}^2_{plants}$ above 0.30 were found for all of the traits except SH (0.19), SN (0.29), SW (0.26), and TCH (0.27) in SR1 and SH (0.23) in SR2. The $\hat{h}^2_{plants}$ ranged from 0.19 (SH) to 0.45 (SD) in SR1 and from 0.23 (SH) to 0.49 (SD) in SR2. Even considering some indicated exceptions, the values show that much of the observed phenotypic variation can be attributed to differences in the genotypic level.

The genotypic and residual coefficients of variation between SR1 and SR2 for each individual trait were similar, with a few exceptions for the residual coefficient of variation (CV_R). The exceptions to the pattern of similarity were as follows: (i) the CV_R for SW was approximately 40% higher in SR1 (25.50) than in SR2 (18.30), and (ii) TCH was approximately 34% higher in SR1 (23.51) than in SR2 (17.54). The values of the estimates of the genetic and phenotypic variances were similar in SR1 and SR2 for each individual trait. An exception was the estimate of the genetic variance component: SN was approximately 77% higher in SR2 (529.90) than in SR1 (298.00). In the estimate of the phenotypic variance component, SW was approximately 70% higher in SR1 (3177.00) than in SR2 (1868.60), TCH was 58% higher in SR1 (2120.00) than in SR2 (1336.00), SD was approximately 53% higher in SR1 (11.10) than in SR2 (7.23), and TPH was 32% higher in SR1 (38.93) than in SR2 (29.46).

The range of variation was different between the SR1 and SR2 families for all of the evaluated traits. Family SR2 showed much higher ranges of variation for the traits of SN, SW, and TCH. The average values of the traits of SH, BRIX, POL%C, POL%J, and FIB showed similar variations between the two evaluated families. However, SD, SN, SW, TPH, and TCH showed differences in the average values between the families. The average values of SD, SW, TCH, and TPH were greater in SR1, whereas the average value of SN was higher in SR2. In SR1, the average of the progeny was higher than the average of both parents for TCH and TPH. In SR2, the average of the progeny was higher than the average of both parents for SD. In addition, in both families and for all traits evaluated, the offspring had higher averages than the parents of the families. These results show the occurrence of transgressive segregation in both families.

Table 4. Ranges, averages, estimates of components of genetic variance ($\hat{\sigma}^2_G$) and phenotype ($\hat{\sigma}^2_P$), coefficients of genotypic variation (CV_G) and residual (CV_R), and broad heritability on a genotype-mean ($\hat{H}^2_{means}$) and individual-plant basis ($\hat{H}^2_{plants}$), stalk diameter (SD) in mm, stalk height (SH) in m, stalk number (SN) by direct counting, stalk weight (SW) in kg, BRIX as °Brix, sucrose content of cane (POL%C) in percentage, sucrose content of juice (POL%J) in percentage, fiber (FIB) as a percentage, cane yield (TCH) in t ha^{-1}, and sucrose yield (TPH) in t ha^{-1} for the two families of sugarcane (SR1 and SR2) at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012, and 2013).

Trait	Range	Avg.	SP80-3280	SP81-3250	RB835486	RB925345	$\hat{\sigma}^2_G$	$\hat{\sigma}^2_P$	CV_G	CV_R	$\hat{H}^2_{means}$	$\hat{H}^2_{plants}$
				SP80-3280 × RB8354B6 (SR1)								
SD, mm	24.18–34.44	28.39	29.57		27.16		4.96	11.10	7.84	8.73	0.92	0.45
SH, m	1.99–2.60	2.31	2.37		2.22		0.01	0.09	5.71	11.65	0.78	0.19
SN	71.80–138.50	105.85	108.10		96.20		298.00	1012.30	16.31	25.24	0.86	0.29
SW, kg	128.40–254.10	184.32	203.50		152.3		822.00	3177.00	15.55	25.50	0.82	0.26
BRIX, °Brix	18.79–22.88	20.85	21.19		21.05		0.63	1.62	3.82	4.76	0.83	0.39
POL%C	13.46–17.26	15.52	15.72		15.84		0.64	1.65	5.15	6.45	0.83	0.39
POL%J	15.92–20.62	18.44	18.63		18.85		0.89	2.27	5.11	6.38	0.83	0.39
FIB, %	10.78–14.59	12.20	12.04		12.34		0.59	1.46	6.30	7.65	0.84	0.40
TCH, t ha^{-1}	114.50–219.90	162.07	95.84		85.48		582.00	2120.00	14.88	23.51	0.83	0.27
TPH, t ha^{-1}	15.87–31.84	23.72	16.00		15.00		13.63	38.93	15.56	21.09	0.81	0.35
				SP81-3250 × RB925345 (SR2)								
SD, mm	21.16–31.64	25.61		25.25		25.18	3.55	7.23	7.35	7.15	0.94	0.49
SH, m	1.98–2.68	2.34		2.25		2.49	0.02	0.08	5.87	10.06	0.83	0.23
SN	50.50–202.40	117.44		127.60		115.20	529.90	1257.60	19.59	22.10	0.92	0.42
SW, kg	64.40–244.30	168.15		172.50		175.70	734.20	1868.60	16.11	18.30	0.90	0.39
BRIX, °Brix	18.27–22.88	20.93		21.02		21.63	0.65	1.59	3.86	4.00	0.81	0.40
POL%C	13.24–17.35	15.54		15.60		16.24	0.53	1.42	4.67	5.42	0.79	0.37
POL%J	15.62–20.77	18.57		18.58		19.47	0.81	2.11	4.84	5.42	0.80	0.38
FIB, %	11.17–15.13	12.64		12.37		12.94	0.52	1.12	5.72	5.66	0.86	0.46
TCH, t ha^{-1}	84.10–214.10	149.04		150.70		157.00	543.10	1336.00	15.63	17.54	0.90	0.40
TPH, t ha^{-1}	14.02–30.93	22.09		22.91		24.26	11.53	29.46	15.36	18.26	0.82	0.39

Genotypic Correlations of Yield Components

Pairwise genotypic correlations among the 10 evaluated traits, considering two locations (Araras and Ipaussu) and three harvests (2011, 2012, and 2013), are shown in Fig. 1. Overall, 17 and 12 significant genotypic correlations ($P < 0.05$) occurred between the evaluated traits in SR1 and SR2, respectively. According to the degree of correlation between the traits, the correlations were grouped into low (≤0.35), moderate (0.36–0.70) and strong (>0.71). Thus, in SR1, seven interactions were classified as low (SD–SN, SD–SW, SD–FIB, SD–TCH, SD–TPH, POL%C–TPH, and POL%J–TPH), three were classified as moderate (SH–SW, SN–SW and SW–TPH), and seven were classified as strong (SN–TCH, SN–TPH, SW–TCH, BRIX–POL%C, BRIX–POL%J, POL%C–POL%J, and TCH–TPH). The correlations SD–SN, SD–FIB, SD–TCH, and SD–TPH were negative (Fig. 1a). In SR2, four interactions were classified as low (SD–SH, SH–FIB, BRIX–FIB, and POL%J–FIB), two were classified as moderate (SD–SN and SD–FIB), and six were classified as strong (SW–TCH, SW–TPH, BRIX–POL%C, BRIX–POL%J, POL%C–POL%J, and TCH–TPH). The correlations SD–SN, SD–FIB, and SH–FIB were negative (Fig. 1b).

Of the total genotypic correlations that were observed, eight were common between SR1 and SR2 (SD–SN, SD–FIB, SW–TCH, SW–TPH, TCH–TPH, BRIX–POL%C, BRIX–POL%J, and POL%C–POL%J). However, three correlations exhibited differences between the two families (SD–SN, SD–FIB, and SW–TPH). The SD–SN and SD–FIB correlations were classified as negative and low in SR1 (–0.31 and –0.29, respectively) and as negative and moderate in SR2 (–0.44 and –0.39, respectively). The SW–TPH correlation was classified as positive and moderate in SR1 (0.62) and positive and strong in SR2 (0.92). The eight exclusive correlations that were present in SR1 were classified as low (SD–SW, SD–TCH, SD–TPH, POL%C–TPH, and POL%J–TPH), moderate (SH–SW and SN–SW) and strong (SN–TCH and SN–TPH), whereas the correlations SD–TCH and SD–TPH were negative (Fig. 1a). In SR2, four exclusive correlations were classified as low (SD–SH, SH–FIB, BRIX–FIB, and POL%J–FIB), whereas the correlation SH–FIB was negative (Fig. 1b).

Probabilities, Segregations, and Heritability of Resistance to Brown Rust

The GLMM-based analysis revealed that approximately 66% (101) and 32% (74) of full-sib genotypes in SR1 and SR2, respectively, have at most a 10% probability of showing symptoms of the disease (Fig. 2), i.e., at least a 90% probability of being resistant under the evaluated local and environmental conditions. In SR1, the parent SP80-3280 showed a 99.50% probability of being resistant to the disease, while the parent RB835486 showed a 99.80% probability of being susceptible to brown rust. In SR2, the parents SP81-3250 and RB925345 showed 88.90% and 96.01% probabilities of being resistant and susceptible to brown rust, respectively (Fig. 2). The segregation that was observed in SR1 showed a strong displacement of the curve toward the class that was considered resistant (up

Agronomy Journal • Volume 108, Issue 5 • 2016

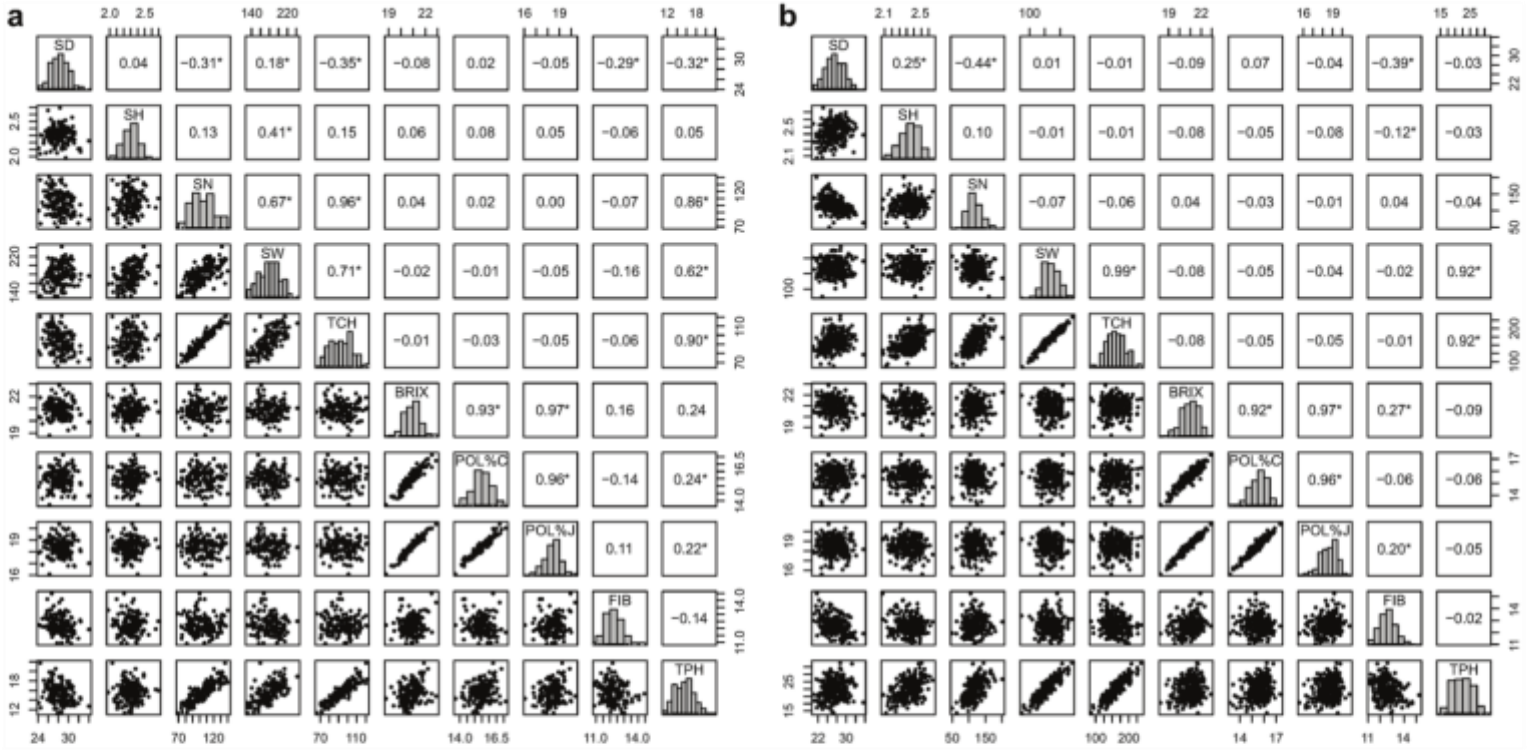


Fig. 1. Estimates of the genotypic correlation of yield components for (a) Family SR1 derived from a cross between SP80-3280 and RB835486, and (b) Family SR2 derived from a cross between SP81-3250 and RB925345 for the stalk diameter (SD) in mm, stalk height (SH) in m, stalk number (SN) by direct counting, stalk weight (SW) in kg, BRIX as °Brix, sucrose content of cane (POL%C) in percentage, sucrose content of juice (POL%J) in percentage, fiber (FIB) as a percentage, cane yield (TCH) in t ha^{-1}, and sucrose yield (TPH) in t ha^{-1} at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012, and 2013). For each trait, the histograms of the adjusted means (diagonal), scatterplots (below diagonal), and values of the genotypic correlation (above diagonal) between pairs of traits are shown. *Significant at the 5% global level ($P < 0.05$).

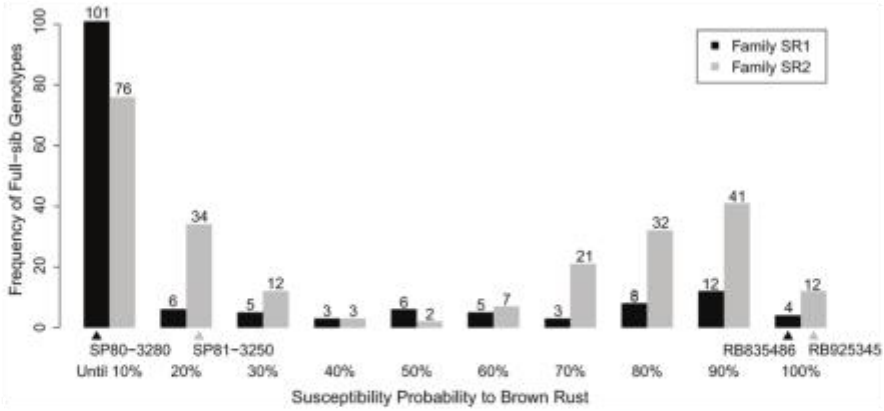


Fig. 2. Frequency of full-sib genotypes in the probability classes of brown rust susceptibility as calculated with a generalized linear mixed model (GLMM) for the two families of sugarcane (SR1, derived from a cross between SP80-3280 and RB835486, and SR2, derived from a cross between SP81-3250 and RB925345) at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012, and 2013).

to a 10% probability of having the disease). In SR2, a bimodal distribution was observed (Fig. 2). The $\hat{H}^2_{\text{plants}}$ of brown rust resistance were high (0.93 and 0.84 in SR1 and SR2, respectively). The genetic coefficient of variation was 8.76 and 5.31 in SR1 and SR2, respectively, and the residual coefficient of variation was 0.62 and 0.95 in SR1 and SR2, respectively.

DISCUSSION

Sugarcane is one of the most important crops worldwide, and its importance is mainly attributed to its derivatives, i.e., sugar and ethanol. Brazil is the world's largest producer of sugarcane (FAO, 2014) and is constantly striving for increased production. However, increasing production requires an adequate and sustainable method for the modernization of sugarcane cultivation. Therefore, high-technology agriculture associated with knowledge of the genetic basis of inheritance of all traits that are of economic interest and are directly linked to sugarcane production is very important for increasing productivity without expanding the area planted with sugarcane.

A set of traits should be simultaneously considered because interest lies in the combined selection of traits rather than isolated traits. Several studies that involved yield components have been conducted on sugarcane, albeit using statistical approaches with a series of limitations (Gallacher, 1997; Lin et al., 1993), specifically, without considering the unbalanced data and variance homogeneity assumptions and without assessing whether there were genetic correlations between harvests and locations for estimating breeding values (Balzarini, 2002; Piepho and Möhring, 2007; Smith et al., 2005). Another important limitation is missing phenotypic data, which is very common in experiments with sugarcane. Modeling these limitations would allow more realistic results that should be easier to apply in a sugarcane genetic breeding program. The mixed models approach is suitable for evaluating the heterogeneity of genetic variances and correlations across environments (Malosetti et al., 2013). Therefore, the use of a more sophisticated statistical model would permit data processing that is more appropriate for the experiment and that signifies a major change in the analysis (i.e., the genotype grouping factors, such as the harvest year and location, would be considered) (Pastina et al., 2012). Although the adjustments for VCOV structures were comprehensively studied for the data in the present study, it is important to note that the use of more locations and harvest years would probably permit the adjustment of other variance and covariance structures. For example, in sugarcane data across harvest years, the data from the same individual with time are expected to be correlated. This result is more evident in families from breeding programs with more harvest years and locations because of the long experimentation process until the release of a new cultivar.

The VCOV matrices were primarily constructed by considering the genetic effects matrix (**G** matrix) and then the nongenetic effects matrix (**R** matrix). In principle, the selection of VCOV models for the location effect requires, among other factors, prior knowledge of the soil and climatic conditions of each evaluated location. For harvests, the biological response of the plants across years is important. Locations (i.e., Araras and Ipaussu) have different soil and climatic conditions, which contribute to complex interactions and possible changes in the responses of genotypes. Harvests experience a drop in productivity with time within one cycle. Thus, the individual models for each measured trait are appropriate, and they can reflect the efficiency and reality of the final genotypic response.

Considering the AIC values for the selection of the best models in the sugarcane families, a preferential selection for SR1 was observed in Models 10 (SD, SW, and TCH) and 12 (BRIX, POL%J, FIB, and TPH) with VCOV structure $\mathbf{G}_p = \mathbf{G}^{\mathrm{L}}_{N\times N} \otimes \mathbf{G}^{\mathrm{H}}_{M\times M}$ (Table 2). Model 10 presents an UNST structure for local and AR1 for harvests, indicating a correlation between successive harvests and a systematic explanation of the existing temporal dependence. The productivity of sugarcane tends to decrease over harvests; therefore, we expect a greater correlation among nearby harvests and a lower correlation among distant harvests due to physiological and genetic changes. Model 12 presents a UNST structure for both locations and harvests; this is a complex model that generally captures all of the possible variations, i.e., the traits that exhibit different variances and covariances between locations

and harvests. Model 5 was selected for SH and SN, and Model 6 was selected for POL%C; these models have a VCOV structure of $\mathbf{G}_P = \mathbf{G}^{\mathrm{E}}_{P\times P}$. Model 6 assumes a general structure $\mathbf{G}_P$ matrix, which is completely unstructured for different genetic variances in each environment and for different covariances between pairs of environments. Model 5 is an approximated unstructured model that can be interpreted as a linear regression model for genotype effects and GEI of environmental covariates, i.e., it measures the sensitivity of the genotype in relation to the "weight" of each environment. This model has been suggested for MET analysis (Burgueño et al., 2012; Kelly et al., 2007; Piepho, 1998; Thompson et al., 2003) because it captures the genetic variation in genotypes in terms of environment as well as the genetic covariance between environments with more realistic and accurate predictions. In SR2, SW was the only trait with selected VCOV $\mathbf{G}_P = \mathbf{G}^{\mathrm{L}}_{N\times N} \otimes \mathbf{G}^{\mathrm{M}}_{M\times M}$ structure (Model 12), and a preferential selection by Model 6 with VCOV $\mathbf{G}_P = \mathbf{G}^{\mathrm{E}}_{P\times P}$ structure was also observed (SD, SH, BRIX, POL%C, POL%J, FIB, TCH, and TPH) (Table 2). The SN and POL%C were the only traits with the same models selected in SR1 and SR2 (Models 5 and 6, respectively) (Table 2). The genetic complexity of sugarcane is also reflected in the response of genotypes under the conditions to which they were submitted and consequently in the model selection that best fit the response pattern of the data. The biomass production of sugarcane is influenced by several factors (genetic, physiological, and environmental), so Models 6 and 12 are most suitable for estimating the genetic parameters of this measure by considering all possible variations. However, this matrix requires an estimation of the maximum number of parameters. In experiments with many locations and harvests, this analysis can become computationally unfeasible. Alternatively, the FA1 matrix, which considers the genetic effects of location (Boer et al., 2007; Burgueño et al., 2012; Kelly et al., 2007; Meyer, 2009; Smith et al., 2007; Thompson et al., 2003), along with the AR1 matrix adjusted for genetic effects of harvests (Pastina et al., 2012), can accurately predict genetic parameters.

Commonly, sugarcane plant breeders independently assess the results of each experiment. This practice is equivalent to the predictions of the DIAG model, which indicates the heterogeneity of variance but not the correlation of performances of genotypes among the experiments (Kelly et al., 2007). Our results show that none of the analyzed traits adjusted to the DIAG model for the $\mathbf{G}_P$ matrix (Table 2). Thus, the model selection approach that adjusts the natural response pattern for each trait is far superior to that currently practiced by plant breeders because it can capture both the heterogeneity of variance and more complex covariance structures at the genetic level, resulting in a more accurate prediction of individual experiments or multiple environments. The implementation of this data analysis model improves the predictive accuracy directly related to the heritability and genetic gain. Sugarcane breeding programs can increase the efficiency of superior genotype selection in every stage of the selection and in diverse environments.

Defined as the heritable portion transmitted to offspring (Falconer and Mackay, 1996), heritability is an important parameter because it determines the response to selection and because it can help select the optimal strategy for a breeding program (Piepho and Möhring, 2007; Sadras et al., 2013). All of the broad hereditability values presented in this study are high (Bernardo, 2010) ranging from 0.78 to 0.92 for SR1 and from 0.79 to 0.94 for SR2; thus, the observed phenotypic variation is mostly due to the genotypic variation in these populations (Table 4). Comparing the results of SR1 and SR2 with those reported in the literature, Hoarau et al. (2002) found lower $\hat{H}^2_{\mathrm{means}}$ values for SD (0.91), SH (0.83), SN (0.86) and BRIX (0.81) in separate populations of selfing of R570 compared with SR1 and SR2, with the exception of SH, which was lower in SR1. Aitken et al. (2006) found $\hat{H}^2_{\mathrm{means}}$ values that were slightly higher for BRIX (0.88) and POL%J (0.93) in separate populations of a cross between Q165 and IJ76-514. Aitken et al. (2008), using the same parents and a population of 227 individuals, found lower $\hat{H}^2_{\mathrm{means}}$ values for SD (0.88), SN (0.83), and TCH (0.71) and higher values for SH (0.85). Pinto et al. (2010), while working with a separate population of a cross between SP80-180 and SP80-4966, found slightly higher $\hat{H}^2_{\mathrm{means}}$ values for POL%C (0.84) and TPH (0.88) and the lowest value for FIB (0.81). The value for TCH (0.87) was greater than that of SR1 and less than that of SR2. Mancini et al. (2012) found lower $\hat{H}^2_{\mathrm{means}}$ values for SD (0.80), SH (0.72), SN (0.76), SW (0.77), BRIX (0.60), POL%C (0.59), FIB (0.75), and TCH (0.70) when evaluating a separate population of a cross between IACSP95-3018 and IACSP93-3046. These studies did not consider the correlations between harvests. The comparison of results revealed that the broad-sense hereditability values were mostly higher in SR1 and SR2. A good experimental control combined with a statistical model that can integrate data at different locations and multiple harvests generates more accurate estimates of heritability.

The range of the phenotypic values for all of the evaluated traits was greater than the phenotypic range of the parents, i.e., transgressive segregation occurred (Table 4) as also observed by Hoarau et al. (2002) and Mancini et al. (2012). Certainly, the selection of checks is a crucial point for the comprehension of the phenotypic values observed in these experiments. In METs, several environments are tested, and each genotype can develop best in a specific environment compared with others. Comparisons with suitable checks are fundamental for efficient BLUP estimation and the identification of the best genotypes. Furthermore, the selection of checks based on considerations of more than one trait is a challenge because it is desirable that all of the check's phenotypic values are within the range of the phenotypic values of the progeny. In our experiments, we used commercial cultivars as checks, i.e., for each family, the parents and a non-parental cultivar were used as checks (Supplemental Table S4). The check non-parental cultivar RB867515 is currently the most cultivated cultivar in Brazil and exhibits high productivity rates in different production environments, i.e., under different types of soil and climatic conditions. The choice of RB867515 as a non-parental check was primarily based on the substantial knowledge of its production behaviors in different environments. Ensuring the best estimates of the phenotypic values via the appropriate choice of checks is a fundamental step that can contribute to breeding programs and result in the release of cultivars with higher yields.

Genotypes can also be selected through genotypic correlation, which combines more than one desirable trait in the same

plant for indirect selection (Ram et al., 1997). Correlations among traits may reflect biological processes that are of considerable evolutionary interest and are the result of genetic, functional, physiological, or developmental nature (Jamoza et al., 2014; Soomro et al., 2006). The common strong genotypic correlation between SR1 and SR2 (SW–TCH, BRIX–POL%C, BRIX–POL%J, POL%C–POL%J, and TCH–TPH) (Fig. 1) shows that the selection practiced by breeding programs has aimed to increase the amount of sugar and the stalk weight, considering that the parents that originated both families had a different genetic background and still gather favorable alleles for the expression of correlated traits. Using these main traits (SW, BRIX, POL%C, POL%J, TCH, and TPH) throughout the selection period of a breeding program, the cultivars may meet the expectations of highly accumulated sucrose and high production in terms of weight. However, the genetic gain for these traits with the conventional breeding process is nearly stagnant (Dal-Bianco et al., 2012). Several limitations inherent in a breeding program may be noted: (i) a lack of knowledge of the genetic material that is present in germplasm banks, which could be used to perform cross-breeding with the greatest potential to generate superior cultivars; (ii) sparse and mismanaged experimental trials; (iii) failures and a lack of standardization in the collection of phenotypic data; (iv) lack of environmental correlation analysis of the phenotypic data; (v) lack of knowledge of the genetic basis of the traits of interest; (vi) neglect of disease and pest occurrence; and (g) low investment in research and biotechnology (Breseghello and Coelho, 2013; Mahon, 1983; Prohens, 2011).

Among the diseases that affect sugarcane, brown rust is present in almost all of the cultivation areas (Asnaghi et al., 2004; Ryan and Egan, 1989). This disease can hinder the performance of cultivars and exclude them from the breeding stock of producing units. Therefore, sugarcane breeding programs should seek sources of resistance and produce cultivars that are able to resist the pathogen. When evaluating the VCOV structures that are adjusted to residues, we found that DIAG was appropriate for SD and SW in SR1 and SW in SR2 for location. In addition, DIAG was appropriate for SD and FIB for the factorial combination between location and harvest in SR2 (Table 3). Thus, residues, which are causes of variation that are not controlled for in these traits, are possibly different or present different intensities between the two locations. Araras and Ipaussu have different soil and climatic characteristics, as discussed above. Moreover, it is important to highlight that Ipaussu has a great natural incidence of brown rust disease due to the fungus *P. melanocephala*, which could strongly interfere with productivity. The extreme compatibility of the fungus that causes brown rust with sugarcane reduces the life of leaves, which lose their photosynthetic function. The presence of sporulating pustules on the leaves results in the reduced growth of sugarcane and significant productivity losses, compromising the final biomass production, depending on the susceptibility of the cultivar and the environmental conditions (Asnaghi et al., 2000; Hoy and Hollier, 2009; McFarlane et al., 2006; Oloriz et al., 2011, 2012; Purdy et al., 1983; Raid and Comstock, 2000; Taylor et al., 1986). The severity of brown rust is assessed on a particular scale (Amorim et al., 1987; Tai et al., 1981), and the frequency of full-sib genotypes in each severity class is influenced by the genetic basis of the parents of the family. Several researchers have reported that brown rust resistance is controlled by one or a few genes (Asnaghi et al., 2004; Costet et al., 2012; Daugrois et al., 1996; Garsmeur et al., 2011; Glynn et al., 2013; Hogarth et al., 1993; Le Cunff et al., 2008; Parco et al., 2014; Raboin et al., 2006; Racedo et al., 2013; Ramdoyal et al., 2000; Sordi et al., 1988). Costet et al. (2012) showed that resistance to brown rust in modern polyploid sugarcane cultivars essentially depends on the major gene *Bru*1. A GLMM was used to analyze data from SR1 and SR2 because the binomial variable resistance–susceptibility was used to characterize the data of the progenies (Bennewitz et al., 2014; Bolker et al., 2009; De Silva et al., 2014). A strong displacement of the curve toward the class that is considered resistant (up to a 10% probability of having the disease) occurred in SR1, and a bimodal distribution was observed for SR2 (Fig. 2). Our results suggest that one or a few genes originating from the parent SP80-3280 and SP81-3250 may transfer resistance to the full-sib genotypes in SR1 and SR2, respectively. The distribution of SR2 was bimodal and allows us to infer that a combination of a different number of copies of the gene conferring resistance or susceptibility to full-sib genotypes from this family is preferred. In addition, because the broad-sense heritability of brown rust resistance had high values of 0.93 and 0.84 in SR1 and SR2, respectively, *Bru*1 may be present and segregating in the progenies.

Therefore, knowledge of the genotypes, the correct orientation of the cross-breeds, experimental trials with highly accurate measurements, the use of models with VCOV matrices (which consider the heterogeneity of the variance and assess the correlations between locations and harvests), and the commitment to generate truly resistant cultivars are fundamental for obtaining more productive sugarcane cultivars. Our results showed that LMMs and GLMMs for estimating genetic parameters in sugarcane are potentially useful in the investigation of the heterogeneous genetic variances and correlations between environments. These models can be tested in other families and METs of sugarcane for an understanding of the complex relationships among traits and environments.

ACKNOWLEDGMENTS

We gratefully acknowledge the Raízen group for their support in the field experiments in Ipaussu. We also acknowledge Luiz Plínio Zavaglia and Sandro Augusto Ferrarez for their support in the field experiments and Rodrigo Gazaffi for his comments on the manuscript. This work was supported by grants from INCT-Bioetanol (Instituto Nacional de Ciência e Tecnologia do Bioetanol), FINEP (Finaciadora de Estudos e Projetos), and FAPESP (Fundação de Amparo à Pesquisa de São Paulo, 08/52197-4). Thiago W.A. Balsalobre, Melina C. Mancini, and Guilherme da S. Pereira received doctorate fellowships from FAPESP (10/50091-4, 10/50549-0 and 12/25236-4, respectively). Carina O. Anoni received a doctorate fellowship from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Fernanda Z. Barreto received a master's fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Anete P. de Souza and Antonio A. F. Garcia received research fellowships from CNPq.

REFERENCES

Aitken, K.S., S. Hermann, K. Karno, G.D. Bonnett, L.C. McIntyre, and P.A. Jackson. 2008. Genetic control of yield related stalk traits in sugarcane. Theor. Appl. Genet. 117:1191–1203. doi:10.1007/s00122-008-0856-6

Aitken, K.S., P.A. Jackson, and C.L. McIntyre. 2006. Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar × *Saccharum officinarum* population. Theor. Appl. Genet. 112:1306–1317. doi:10.1007/s00122-006-0233-2

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19:716–723. doi:10.1109/TAC.1974.1100705

Amorim, L., A. Bergamin-filho, A. Sanguino, C. Cardoso, V.A. Moraes, and C.R. Fernandes. 1987. Metodologia de avaliação da ferrugem da cana-de-açúcar (*Puccinia melanocephala*). Bol. Tec. COPERSUCAR 39:13–16.

Asnaghi, C., F. Paulet, C. Kaye, L. Grivet, J.C. Glaszmann, and A. D'Hont. 2000. Application of synteny across the Poaceae to determine the map location of a rust resistance gene of sugarcane. Theor. Appl. Genet. 101:962–969. doi:10.1007/s001220051568

Asnaghi, C., D. Roques, S. Ruffel, C. Kaye, J.Y. Hoarau, H. Télismart, et al. 2004. Targeted mapping of a sugarcane rust resistance gene (*Bru*1) using bulked segregant analysis and AFLP markers. Theor. Appl. Genet. 108:759–764. doi:10.1007/s00122-003-1487-6

Balzarini, M. 2002. Applications of mixed models in plant breeding. In: M.S. Kang, editor, Quantitative genetics, genomics and plant breeding. CABI Publ., New York. p. 353–363.

Beaulieu, J., T. Doerksen, S. Clément, J. Mackay, and J. Bousquet. 2014. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity 113:343–352. doi:10.1038/hdy.2014.36

Bennewitz, J., S. Bögelein, P. Stratz, M. Rodehutscord, H.P. Piepho, J.B. Kjaer, and W. Bessei. 2014. Genetic parameters for feather pecking and aggressive behavior in a large F_2–cross of laying hens using generalized linear mixed models. Poult. Sci. 93:810–817. doi:10.3382/ps.2013-03638

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury, MN.

Bevan, M.W., and C. Uauy. 2013. Genomics reveals new landscapes for crop improvement. Genome Biol. 14:206. doi:10.1186/gb-2013-14-6-206

Boer, M.P., D. Wright, L. Feng, D.W. Podlich, L. Luo, M. Cooper, and F.A. van Eeuwijk. 2007. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177:1801–1813. doi:10.1534/genetics.107.071068

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens, and J.-S.S. White. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. Trends Ecol. Evol. 24:127–135. doi:10.1016/j.tree.2008.10.008

Breseghello, F., and A.S. Coelho. 2013. Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.). J. Agric. Food Chem. 61:8277–8286. doi:10.1021/jf305531j

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52:707–719. doi:10.2135/cropsci2011.06.0299

Costet, L., L. Le Cunff, S. Royaert, L.M. Raboin, C. Hervouet, L. Toubi, et al. 2012. Haplotype structure around *Bru*1 reveals a narrow genetic basis for brown rust resistance in modern sugarcane cultivars. Theor. Appl. Genet. 125:825–836. doi:10.1007/s00122-012-1875-x

Cox, M., D. Hogarth, and P. Hansen. 1994. Breeding and selection for high early season sugar content in a sugarcane (*Saccharum* spp. hybrids) improvement program. Aust. J. Agric. Res. 45:1569–1575. doi:10.1071/AR9941569

Crossa, J., Y. Beyene, S. Kassa, P. Pérez, J.M. Hickey, C. Chen et al. 2013. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 3:1903–1926. doi:10.1534/g3.113.008227

Dal-Bianco, M., M.S. Carneiro, C.T. Hotta, R.G. Chapola, H.P. Hoffmann, A.A. Garcia, and G.M. Souza. 2012. Sugarcane improvement: How far can we go? Curr. Opin. Biotechnol. 23:265–270. doi:10.1016/j.copbio.2011.09.002

Daugrois, J.H., L. Grivet, D. Roques, J.Y. Hoarau, H. Lombard, J.C. Glaszmann, and A. D'Hont. 1996. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570.' Theor. Appl. Genet. 92:1059–1064. doi:10.1007/BF00224049

De Silva, N.H., L. Gea, and R. Lowe. 2014. Genetic analysis of resistance to *Pseudomonas syringae pv. actinidiae* (Psa) in a kiwifruit progeny test: An application of generalised linear mixed models (GLMMs). SpringerPlus 3:547. doi:10.1186/2193-1801-3-547

de Vries, S.C., G.W.J. van de Ven, M.K. van Ittersum, and K.E. Giller. 2010. Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. Biomass Bioenergy 34:588–601. doi:10.1016/j.biombioe.2010.01.001

D'Hont, A. 2005. Unravelling the genome structure of polyploids using FISH and GISH: Examples in sugarcane and banana. Cytogenet. Genome Res. 109:27–33. doi:10.1159/000082378

D'Hont, A., and J.C. Glaszmann. 2001. Sugarcane genome analysis with molecular markers: A first decade research. Proc. Int. Soc. Sugar-Cane Technol. 24:556–559.

D'Hont, A., D. Ison, K. Alix, C. Roux, and J.C. Glaszmann. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41:221–225. doi:10.1139/g98-023

Edwards, D., J. Batley, and R.J. Snowdon. 2013. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 126:1–11. doi:10.1007/s00122-012-1964-x

Eksteen, A., A. Singels, and S. Ngxaliwe. 2014. Water relations of two contrasting sugarcane genotypes. Field Crops Res. 168:86–100. doi:10.1016/j.fcr.2014.08.008

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Harlow, UK.

FAO. 2014. FAOSTAT. [Database.] FAO, Rome. http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#anco (accessed 10 Nov. 2014).

Gallacher, D.J. 1997. Evaluation of sugarcane morphological descriptors using variance component analysis. Aust. J. Agric. Res. 48:769–774. doi:10.1071/A96062

Garsmeur, O., C. Charron, S. Bocs, V. Jouffe, S. Samain, A. Couloux, et al. 2011. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. New Phytol. 189:629–642. doi:10.1111/j.1469-8137.2010.03497.x

Gerbens-Leenes, W., A.Y. Hoekstra, and T.H. van der Meer. 2009. The water footprint of bioenergy. Proc. Natl. Acad. Sci. 106:10219–10223. doi:10.1073/pnas.0812619106

Gianola, D., and J. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201–224. doi:10.1186/1297-9686-15-2-201

Glynn, N.C., C. Laborde, R.W. Davidson, M.S. Irey, B. Glaz, A. D'Hont, and J.C. Comstock. 2013. Utilization of a major brown rust resistance gene in sugarcane breeding. Mol. Breed. 31:323–331. doi:10.1007/s11032-012-9792-x

Goldemberg, J. 2007. Ethanol for a sustainable energy future. Science 315:808–810. doi:10.1126/science.1137013

Gouy, M., D. Luquet, L. Rouan, J.-F. Martiné, A. Thong-Chane, L. Costet, et al. 2015. Site and *Saccharum spontaneum* introgression level drive sugarcane yield component traits and their impact on sucrose yield in contrasted radiation and thermal conditions in La Réunion. Field Crops Res. 171:99–108. doi:10.1016/j.fcr.2014.11.002

Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal, D. Roques, et al. 2013. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126:2575–2586. doi:10.1007/s00122-013-2156-z

Grivet, L., and P. Arruda. 2002. Sugarcane genomics: Depicting the complex genome of an important tropical crop. Curr. Opin. Plant Biol. 5:122–127. doi:10.1016/S1369-5266(02)00234-0

Ha, S., P.H. Moore, D. Heinz, S. Kato, N. Ohmido, and K. Fukui. 1999. Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescence in situ hybridization and imaging methods. Plant Mol. Biol. 39:1165–1173. doi:10.1023/A:1006133804170

Henderson, C.R. 1984. Applications of linear models in animal breeding. Univ. of Guelph, Guelph, ON, Canada.

Hoarau, J.-Y., L. Grivet, B. Offmann, L.M. Raboin, J.P. Diorflar, J. Payet, et al. 2002. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.): II. Detection of QTLs for yield components. Theor. Appl. Genet. 105:1027–1037. doi:10.1007/s00122-002-1047-5

Hogarth, D. 1971. Quantitative inheritance studies in sugar-cane: II. Correlations and predicted responses to selection. Aust. J. Agric. Res. 22:103–109. doi:10.1071/AR9710103

Hogarth, D.M., C.C. Ryan, and P.W.J. Taylor. 1993. Quantitative inheritance of rust resistance in sugarcane. Field Crops Res. 34:187–193. doi:10.1016/0378-4290(93)90006-9

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9–113.

Hoy, J.W., and C.A. Hollier. 2009. Effect of brown rust on yield of sugarcane in Louisiana. Plant Dis. 93:1171–1174. doi:10.1094/PDIS-93-11-1171

Irvine, J.E. 1999. *Saccharum* species as horticultural classes. Theor. Appl. Genet. 98:186–194. doi:10.1007/s001220051057

Jackson, P.A., D. Horsley, J. Foreman, D.M. Hogarth, and A.W. Wood. 1991. Genotype × environment (GE) interaction in sugarcane variety trials in the Herbert. Proc. Aust. Soc. Sugar Cane Technol. 13:103–109.

Jamoza, J.E., J. Owuoche, O. Kiplagat and W. Opile. 2014. Broad-sense heritability estimation and correlation among sugarcane (*Saccharum* spp. hybrids) yield and some agronomic traits in western Kenya. Int. J. Agric. Policy Res. 2:016–025.

Kelly, A.M., A.B. Smith, J.A. Eccleston, and B.R. Cullis. 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci. 47:1063–1070. doi:10.2135/cropsci2006.08.0540

Le Cunff, L., O. Garsmeur, L.M. Raboin, J. Pauquet, H. Telismart, A. Selvi, et al. 2008. Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (*Bru*1) in highly polyploid sugarcane ($2n \sim 12x \sim 115$). Genetics 180:649–660. doi:10.1534/genetics.108.091355

Lin, J.F., R.K. Chen, and Y.Q. Lin. 1993. The inheritance of sugar characters in sugarcane. J. Fujian Agric. Coll. 22:392–397.

Macrelli, S., M. Galbe, and O. Wallberg. 2014. Effects of production and market factors on ethanol profitability for an integrated first and second generation ethanol plant using the whole sugarcane as feedstock. Biotechnol. Biofuels 7:26–42. doi:10.1186/1754-6834-7-26

Mahon, J.D. 1983. Limitations to the use of physiological variability in plant breeding. Can. J. Plant Sci. 63:11–21. doi:10.4141/cjps83-002

Malosetti, M., J.M. Ribaut, and F.A. van Eeuwijk. 2013. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4:44. doi:10.3389/fphys.2013.00044

Mancini, M.C., D.C. Leite, D. Perecin, M.A.P. Bidóia, M.A. Xavier, M.G.A. Landell, and L.R. Pinto. 2012. Characterization of the genetic variability of a sugarcane commercial cross through yield components and quality parameters. Sugar Tech 14:119–125. doi:10.1007/s12355-012-0141-5

Matsuoka, S., A.J. Kennedy, E.G.D. dos Santos, A.L. Tomazela, and L.C.S. Rubio. 2014. Energy cane: Its concept, development, characteristics, and prospects. Adv. Bot. 2014:597275 . doi:10.1155/2014/597275

McFarlane, K., S.A. McFarlane, D. Moodley, and R.S. Rutherford. 2006. Fungicide trials to determine the effect of brown rust on the yield of sugarcane variety N29. Proc. S. Afr. Sugar Technol. Assoc. 79:297–300.

Meyer, K. 2009. Factor-analytic models for genotype × environment type problems and structured covariance matrices. Genet. Sel. Evol. 41:21. doi:10.1186/1297-9686-41-21

Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. doi:10.1111/j.1439-0388.2007.00700.x

Naik, S.N., V.V. Goud, P.K. Rout, and A.K. Dalai. 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. 14:578–597. doi:10.1016/j.rser.2009.10.003

Oloriz, M.I., V. Gil, L. Rojas, O. Portal, Y. Izquierdo, E. Jiménez, and M. Hofte. 2012. Sugarcane genes differentially expressed in response to *Puccinia melanocephala* infection: Identification and transcript profiling. Plant Cell Rep. 31:955–969. doi:10.1007/s00299-011-1216-6

Oloriz, M.I., V. Gil, L. Rojas, N. Veitía, M. Höfte, and E. Jiménez. 2011. Selection and characterisation of sugarcane mutants with improved resistance to brown rust obtained by induced mutation. Crop Pasture Sci. 62:1037–1044. doi:10.1071/CP11180

Parco, A.S., M.C. Avellaneda, A.H. Hale, J.W. Hoy, M.J. Pontif, K.A. Gravois, et al. 2014. Frequency and distribution of the brown rust resistance gene *Bru*1 and implications for the Louisiana sugarcane breeding programme. Plant Breed. 133:654–659. doi:10.1111/pbr.12186

Pastina, M.M., M. Malosetti, R. Gazaffi, M. Mollinari, G.R. Margarido, K.M. Oliveira, et al. 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor. Appl. Genet. 124:835–849. doi:10.1007/s00122-011-1748-8

Payne, R.W., D.A. Murray, S.A. Harding, D.B. Baird, and D.M. Soutar. 2009. GenStat for Windows. 12th ed. VSN Int., Hemel Hempstead, UK.

Piepho, H. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance–covariance structures. Theor. Appl. Genet. 97:195–201. doi:10.1007/s001220050885

Piepho, H.P., and J. Möhring. 2007. Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881–1888. doi:10.1534/genetics.107.074229

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. doi:10.1007/s10681-007-9449-8

Pinto, L.R., A.A.F. Garcia, M.M. Pastina, L.H.M. Teixeira, J.A. Bressiani, E.C. Ulian, et al. 2010. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. Euphytica 172:313–327. doi:10.1007/s10681-009-9988-2

Piperidis, G., N. Piperidis, and A. D'Hont. 2010. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol. Genet. Genomics 284:65–73. doi:10.1007/s00438-010-0546-3

Prohens, J. 2011. Plant breeding: A success story to be continued thanks to the advances in genomics. Front. Plant Sci. 2:51. doi:10.3389/fpls.2011.00051

Purdy, L.H., L.J. Liu, and J.L. Dean. 1983. Sugarcane rust, a newly important disease. Plant Dis. 67:1292–1296. doi:10.1094/PD-67-1292

Raboin, L.M., K.M. Oliveira, L. Lecunff, H. Telismart, D. Roques, M. Butterfield, et al. 2006. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: Identification of a gene controlling stalk colour and a new rust resistance gene. Theor. Appl. Genet. 112:1382–1391. doi:10.1007/s00122-006-0240-3

Racedo, J., M.F. Perera, R. Bertani, C. Funes, V. González, M.I. Cuenya, et al. 2013. *Bru*1 gene and potential alternative sources of resistance to sugarcane brown rust disease. Euphytica 191:429–436. doi:10.1007/s10681-013-0905-3

Raid, R.N., and J.C. Comstock. 2000. Common rust. In: P. Rott et al., editors, A guide to sugarcane diseases. CIRAD and ISSCT, Montpellier, France. p. 85–89.

Ram, B., B.S. Chaudhary, and D.K. Yadav. 1997. General and specific selection indices for single stool stages of selection in sugarcane. Euphytica 95:39–44. doi:10.1023/A:1002965924609

Ramburan, S. 2014. A multivariate illustration and interpretation of non-repeatable genotype × environment interactions in sugarcane. Field Crops Res. 157:57–64. doi:10.1016/j.fcr.2013.12.009

Ramburan, S., M. Zhou, and M.T. Labuschagne. 2012. Investigating test site similarity, trait relations and causes of genotype × environment interactions of sugarcane in the midlands region of South Africa. Field Crops Res. 129:71–80. doi:10.1016/j.fcr.2012.01.017

Ramdoyal, K., S. Sullivan, L.C.Y. Lim Shin Chong, G.H. Badaloo, S. Sautally, and R. Domaingue. 2000. The genetics of rust resistance in sugarcane seedling populations. Theor. Appl. Genet. 100:557–563. doi:10.1007/s001220050073

Revelle, W. 2014. PSYCH: Procedures for psychological, psychometric, and personality research. R Found. Stat. Comput., Vienna. http://cran.r-project.org/package=psych (accessed 12 Nov. 2013).

Ryan, C.C., and B.T. Egan. 1989. Rust. In: C. Ricaud and B.T. Egan, editors, Diseases of sugarcane: Major diseases. Elsevier Science Publ., Amsterdam. p. 189–210.

Sadras, V.O., G.J. Rebetzke, and G.O. Edmeades. 2013. The phenotype and the components of phenotypic variance of crop traits. Field Crops Res. 154:255–259. doi:10.1016/j.fcr.2013.10.001

Saini, J.K., R. Saini, L. Tewari. 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. 3 Biotech. 5:337–353. doi:10.1007/s13205-014-0246-5

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464. doi:10.1214/aos/1176344136

Searle, S.R., G. Casella, and C.E. McCulloch. 1992. Variance components. John Wiley & Sons, New York.

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143:449–462. doi:10.1017/S0021859605005587

Smith, A.B., J.K. Stringer, X. Wei, and B.R. Cullis. 2007. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. Euphytica 157:253–266. doi:10.1007/s10681-007-9418-2

Soomro, A.F., S. Junejo, A. Ahmed, and M. Aslam. 2006. Evaluation of different promising sugarcane varieties for some quantitative and qualitative attributes under Thatta (Pakistan) conditions. Int. J. Agric. Biol. 8:195–197.

Sordi, R.A., H. Arizono, and S. Matsuoka. 1988. Indicadores de herdabilidade e avaliação da resistência de clones RB à ferrugem da cana-de-açúcar. Bras. Acucareiro 106:18–23.

State of Sao Paulo Sugarcane, Sugar and Alcohol Growers Council. 2006. Manual de instruções. 5th ed. CONSECANA-SP, Piracicaba, SP, Brazil.

Tai, P.Y.P., J.D. Miller, and J.L. Dean. 1981. Inheritance of resistance to rust in sugarcane. Field Crops Res. 4:261–268. doi:10.1016/0378-4290(81)90077-0

Taylor, P.W.J., B.J. Croft, and C.C. Ryan. 1986. Studies into the effect of sugarcane rust (*Puccinia melanocephala*) on yield. Proc. Int. Soc. Sugar-Cane Technol. 9:411–419.

Thompson, R. 1979. Sire evaluation. Biometrics 35:339–353. doi:10.2307/2529955

Thompson, R., B. Cullis, A. Smith, and A. Gilmour. 2003. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. Aust. N.Z. J. Stat. 45:445–459. doi:10.1111/1467-842X.00297

Trust, B. 2014. Estimation in generalized linear models with random effects. Biometrika 78:719–727.

Waclawovsky, A.J., P.M. Sato, C.G. Lembke, P.H. Moore, and G.M. Souza. 2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. Plant Biotechnol. J. 8:263–276. doi:10.1111/j.1467-7652.2009.00491.x

Welham, S.J., B.J. Gogel, A.B. Smith, R. Thompson, and B.R. Cullis. 2010. A comparison of analysis methods for late-stage variety evaluation trials. Aust. N.Z. J. Stat. 52:125–149. doi:10.1111/j.1467-842X.2010.00570.x

Wolc, A., C. Stricker, J. Arango, P. Settar, J.E. Fulton, N.P. O'Sullivan, et al. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. Genet. Sel. Evol. 43:5. doi:10.1186/1297-9686-43-5

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Supplementary Material

Supplemental Table S1 - Selected models for the $\mathbf{G_p}$ matrix considering each trait separately. The AIC and BIC criteria were used to compare the structures of the variance-covariance matrix. The models for the the $\mathbf{G_p}$ matrix were selected according to the lowest value of the AIC criterion for the SD, SH, SN, SW, BRIX, POL%C, POL%J, FIB, TCH and TPH for the two families of sugarcane (SR1 and SR2) at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012 and 2013).

			SP80-3280 x RB835486 (SR1)			SP81-3250 x RB925345 (SR2)		
Trait	$\mathbf{G}_p$ matrix	Model	n_{PAR}	AIC	BIC	n_{PAR}	AIC	BIC
SD	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	19464.6	19483.4	1	21361.2	21380.6
		(2) UNIF	-	NA	NA	2	20266.0	20291.9
		(3) DIAG	5	19457.5	19501.5	5	21334.3	21386.0
		(4) CS_{Het}	-	NA	NA	6	20155.6	20213.7
		(5) FA1	-	NA	NA	-	NA	NA
		(6) UNST	-	NA	NA	10	**20096.9**	20245.3
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	19226.2	19257.7	4	21361.2	21380.6
		(8) UNST ⊗ UNIF	6	19219.3	19263.3	5	20239.3	20278.1
		(9) UNST ⊗ DIAG	-	NA	NA	6	20794.3	20839.5
		(10) UNST ⊗ AR1	5	**18930.4**	**18968.1**	5	20227.6	20266.3
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	20148.9	**20200.5**
		(12) UNST ⊗ UNST	9	18949.4	19012.2	6	20141.1	20205.7
SH	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	1507.0	1525.9	1	1226.5	1245.9
		(2) UNIF	2	1339.3	1364.4	2	824.9	850.7
		(3) DIAG	5	1430.5	1474.5	5	1227.0	1278.7
		(4) CS_{Het}	6	1285.1	**1335.4**	6	824.5	882.6
		(5) FA1	10	**1274.4**	1349.8	10	796.9	887.3
		(6) UNST	-	NA	NA	10	**728.2**	876.7
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	1361.1	1392.5	4	1086.5	1118.8
		(8) UNST ⊗ UNIF	5	1301.1	1338.8	5	796.8	835.5
		(9) UNST ⊗ DIAG	6	1347.0	1391.0	6	1085.7	1130.9
		(10) UNST ⊗ AR1	5	1298.3	1336.0	5	790.0	**828.7**
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	792.0	843.7
		(12) UNST ⊗ UNST	-	NA	NA	6	789.0	853.6
SN	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	38188.9	38207.7	1	46479.0	46498.4
		(2) UNIF	-	NA	NA	2	45579.2	45605.1
		(3) DIAG	5	38137.5	38181.5	5	46390.4	46442.2
		(4) CS_{Het}	-	NA	NA	6	45462.0	45520.1
		(5) FA1	10	**37808.5**	**37884.0**	10	**45429.0**	45519.6
		(6) UNST	-	NA	NA	-	NA	NA
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	38047.1	38078.5	4	45993.9	46026.2
		(8) UNST ⊗ UNIF	-	NA	NA	5	45465.0	45503.8
		(9) UNST ⊗ DIAG	6	38046.0	38090.0	6	45987.2	46032.5
		(10) UNST ⊗ AR1	-	NA	NA	5	45459.7	45498.5
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	45439.3	**45491.0**
		(12) UNST ⊗ UNST	-	NA	NA	6	45431.5	45496.2

Supplemental Table S1 - Continued.

Trait	$\mathbf{G}_p$ matrix	Model	SP80-3280 x RB835486 (SR1) n_{PAR}	AIC	BIC	SP81-3250 x RB925345 (SR2) n_{PAR}	AIC	BIC
SW	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	42807.7	42826.5	1	48123.3	48142.7
		(2) UNIF	2	42597.3	42622.4	2	47401.2	47427.0
		(3) DIAG	5	42699.3	42743.4	5	47954.3	48006.0
		(4) CS_{Het}	6	42503.2	42553.5	6	47153.9	47212.1
		(5) FA1	-	NA	NA	-	NA	NA
		(6) UNST	-	NA	NA	-	NA	NA
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	42623.0	42654.5	4	47691.2	47723.5
		(8) UNST ⊗ UNIF	5	42504.9	42542.6	5	47093.4	47132.2
		(9) UNST ⊗ DIAG	6	42615.1	42659.1	6	47680.3	47725.6
		(10) UNST ⊗ AR1	5	**42491.7**	**42529.4**	5	47073.1	47111.9
		(11) UNST ⊗ CS_{Het}	7	42522.6	42572.9	7	47063.4	47115.1
		(12) UNST ⊗ UNST	-	NA	NA	6	**47036.9**	**47101.5**
BRIX	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	6657.6	6674.6	1	8244.7	8262.3
		(2) UNIF	-	NA	NA	2	7818.5	**7842.0**
		(3) DIAG	4	6661.5	6695.5	5	8229.5	8270.7
		(4) CS_{Het}	-	NA	NA	6	7801.8	7848.8
		(5) FA1	-	NA	NA	10	7774.8	7845.3
		(6) UNST	10	6439.5	6507.4	10	**7767.2**	7867.1
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	6582.0	6610.3	4	8137.6	8167.0
		(8) UNST ⊗ UNIF	-	NA	NA	5	7821.5	7856.8
		(9) UNST ⊗ DIAG	5	6583.5	6617.4	6	8128.2	8169.3
		(10) UNST ⊗ AR1	5	6445.2	6479.2	5	7816.0	7851.3
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	7808.8	7855.8
		(12) UNST ⊗ UNST	6	**6439.5**	**6479.2**	6	7804.1	7862.9
POL%C	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	6652.7	6669.7	1	8065.8	8083.5
		(2) UNIF	-	NA	NA	2	7691.9	**7715.4**
		(3) DIAG	4	6644.6	6678.6	5	8046.7	8087.9
		(4) CS_{Het}	-	NA	NA	6	7672.9	7719.9
		(5) FA1	8	6418.2	6474.8	10	7663.4	7733.9
		(6) UNST	10	**6407.1**	6475.0	10	**7653.2**	7753.0
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	6554.0	6582.3	4	7983.6	8013.0
		(8) UNST ⊗ UNIF	-	NA	NA	5	7692.0	7727.3
		(9) UNST ⊗ DIAG	5	6555.3	6589.3	6	7980.3	8021.4
		(10) UNST ⊗ AR1	5	6418.5	6452.5	5	7691.1	7726.3
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	7687.3	7734.3
		(12) UNST ⊗ UNST	6	**6410.2**	**6449.8**	6	7688.4	7747.2
POL%J	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	7348.4	7365.4	1	9063.3	9080.9
		(2) UNIF	-	NA	NA	2	8671.5	**8695.0**
		(3) DIAG	4	7348.0	7382.0	5	9048.9	9090.0
		(4) CS_{Het}	-	NA	NA	6	8656.3	8656.3
		(5) FA1	-	NA	NA	10	8644.6	8715.1
		(6) UNST	10	7110.7	7178.7	10	**8637.5**	8737.4
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	7261.8	7290.1	4	8976.1	9005.5
		(8) UNST ⊗ UNIF	-	NA	NA	5	8675.5	8710.7
		(9) UNST ⊗ DIAG	5	7263.6	7297.6	6	11221.6	9013.6
		(10) UNST ⊗ AR1	5	7119.3	7153.2	5	8673.6	8708.8
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	8669.1	8716.1
		(12) UNST ⊗ UNST	6	**7109.1**	**7148.8**	6	8668.9	8727.6

Supplemental Table S1 - Continued.

Trait	$\mathbf{G}_p$ matrix	Model	SP80-3280 x RB835486 (SR1)			SP81-3250 x RB925345 (SR2)		
			n_{PAR}	AIC	BIC	n_{PAR}	AIC	BIC
FIB	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	6325.7	6342.7	1	7275.9	7293.5
		(2) UNIF	-	NA	NA	2	6695.4	6718.9
		(3) DIAG	4	6311.7	6345.6	5	7273.2	7314.3
		(4) CS_{Het}	-	NA	NA	6	6685.3	6732.3
		(5) FA1	8	6059.5	6116.2	-	NA	NA
		(6) UNST	-	NA	NA	10	**6630.1**	6730.0
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	6192.9	6221.2	4	7104.2	7133.6
		(8) UNST ⊗ UNIF	-	NA	NA	5	6685.9	6721.1
		(9) UNST ⊗ DIAG	5	6190.3	6224.3	6	7107.5	7148.7
		(10) UNST ⊗ AR1	5	6069.5	6103.5	5	6683.5	**6718.8**
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	6687.3	6734.3
		(12) UNST ⊗ UNST	6	**6050.5**	**6090.1**	6	6681.0	6739.7
TCH	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	41149.3	41168.1	1	46315.5	46334.9
		(2) UNIF	2	40915.5	40940.6	2	45538.9	45564.7
		(3) DIAG	5	41127.6	41171.6	5	46295.4	46347.1
		(4) CS_{Het}	6	40902.1	40952.4	6	45507.7	45565.8
		(5) FA1	-	NA	NA	10	45489.9	45580.3
		(6) UNST	-	NA	NA	10	**45244.1**	45392.7
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	36746.6	36778.0	4	46047.4	46079.7
		(8) UNST ⊗ UNIF	-	NA	NA	5	45375.0	45413.7
		(9) UNST ⊗ DIAG	6	36749.6	36793.6	6	46037.0	46082.2
		(10) UNST ⊗ AR1	5	**36511.5**	**36549.2**	5	45357.9	45396.7
		(11) UNST ⊗ CS_{Het}	-	NA	NA	7	45354.4	45406.1
		(12) UNST ⊗ UNST	-	NA	NA	6	45321.1	**45385.7**
TPH	$\mathbf{G}_p = \mathbf{G}^{\mathbf{E}}_{P\times P}$	(1) ID	1	13445.7	13462.7	1	15986.7	16004.3
		(2) UNIF	2	13271.2	13293.9	2	15542.8	15566.3
		(3) DIAG	4	13443.0	13477.0	5	15975.9	16022.9
		(4) CS_{Het}	5	13268.4	13308.1	6	15522.1	15574.9
		(5) FA1	8	13271.8	13328.5	10	15508.3	15590.5
		(6) UNST	10	13250.7	13318.7	10	**15411.7**	15546.7
	$\mathbf{G}_p = \mathbf{G}^{\mathbf{L}}_{M\times M}\otimes\mathbf{G}^{\mathbf{H}}_{N\times N}$	(7) UNST ⊗ ID	4	11782.0	11810.3	4	15875.7	15905.1
		(8) UNST ⊗ UNIF	-	NA	NA	-	NA	NA
		(9) UNST ⊗ DIAG	5	11784.0	11818.0	6	15877.8	15918.9
		(10) UNST ⊗ AR1	-	NA	NA	5	15468.7	**15504.0**
		(11) UNST ⊗ CS_{Het}	-	NA	NA	-	NA	NA
		(12) UNST ⊗ UNST	6	**11646.2**	**11685.83**	6	15464.7	15523.4

Supplemental Table S2 - Selected models for the **R** matrix considering each trait separately. The models for the **R** matrix were selected according to the lowest value of the BIC criterion for the SD, SH, SN, SW, BRIX, POL%C, POL%J, FIB, TCH and TPH for family of sugarcane SR1 at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012 and 2013).

			SP80-3280 x RB835486 (SR1)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{IJ}$)	Model	n_{PAR}	AIC	BIC
SD	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	18930.4	18968.1
		DIAG	2	18906.2	**18950.1**
		UNST	3	18907.1	18957.4
	$\mathbf{R}^{H}_{M\times M}$	ID	1	18906.2	18950.1
		DIAG	2	18869.0	18925.5
		UNIF	3	18757.5	18807.7
		CS_{Het}	4	18716.2	**18779.1**
		AR1	2	18823.1	18873.4
		UNST	6	18715.6	18791.0
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	18716.2	18779.1
		DIAG	3	18653.8	**18729.2**
		UNIF	2	18717.1	18786.2
		CS_{Het}	4	18654.7	18736.4
		UNST	6	18658.4	18752.6
SH	$\mathbf{R}_P = \mathbf{R}^{E}_{P\times P}$	ID	1	1274.5	1349.9
		DIAG	-	NA	NA
		UNIF	2	1266.1	**1347.8**
		CS_{Het}	-	NA	NA
		FA1	-	NA	NA
		UNST	-	NA	NA
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	1266.1	**1347.8**
		DIAG	-	NA	NA
		UNIF	2	1267.1	1355.1
		CS_{Het}	-	NA	NA
		UNST	-	NA	NA
SN	$\mathbf{R}_P = \mathbf{R}^{E}_{P\times P}$	ID	1	37808.6	37884.0
		DIAG	-	NA	NA
		UNIF	2	37534.3	**37616.0**
		CS_{Het}	-	NA	NA
		FA1	-	NA	NA
		UNST	-	NA	NA
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	37534.3	**37616.0**
		DIAG	-	NA	NA
		UNIF	2	37529.0	37617.0
		CS_{Het}	-	NA	NA
		UNST	-	NA	NA

Supplemental Table S2 - Continued.

			SP80-3280 x RB835486 (SR1)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{LI}$)	Model	n_{PAR}	AIC	BIC
SW	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	42491.7	42529.4
		DIAG	2	41985.3	**42029.3**
		UNST	3	41986.7	42037.0
	$\mathbf{R}^{H}_{M\times M}$	ID	1	41985.3	42029.3
		DIAG	3	41868.9	41925.5
		UNIF	2	41786.4	41836.7
		CS_{Het}	4	41651.0	41713.9
		AR1	2	41808.9	41859.2
		UNST	6	41608.4	**41683.9**
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	41608.4	41683.9
		DIAG	3	41388.9	41476.9
		UNIF	2	41592.5	41674.2
		CS_{Het}	4	41371.4	**41465.7**
		UNST	6	41370.9	41477.8
BRIX	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	6439.5	**6479.2**
		DIAG	2	6438.9	6484.2
		UNST	3	6439.7	6490.7
	$\mathbf{R}^{H}_{M\times M}$	ID	1	6439.5	6479.2
		DIAG	2	6428.8	6474.1
		UNIF	2	6335.9	6381.2
		CS_{Het}	-	NA	NA
		AR1	-	NA	NA
		UNST	3	6323.4	**6374.4**
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	6428.8	**6474.1**
		DIAG	2	6430.8	6481.8
		UNIF	-	NA	NA
		CS_{Het}	-	NA	NA
		UNST	3	6432.8	6489.4
POL%C	$\mathbf{R}_P = \mathbf{R}^{E}_{P\times P}$	ID	1	6407.1	6475.0
		DIAG	4	6380.7	6465.7
		UNIF	2	6388.1	6461.7
		CS_{Het}	5	6360.7	6451.3
		FA1	8	6342.0	6449.7
		UNST	10	6304.7	**6423.6**
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	6304.7	**6423.6**
		DIAG	2	6306.2	6430.8
		UNIF	2	6306.7	6431.3
		CS_{Het}	3	6308.2	6438.5
		UNST	3	6308.2	6438.5

Supplemental Table S2 - Continued.

			SP80-3280 x RB835486 (SR1)		
Trait	$\mathbf{R}$ matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{IJ}$)	Model	n_{PAR}	AIC	BIC
POL%J	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	7109.1	**7148.8**
		DIAG	2	7110.6	7155.9
		UNST	3	7112.2	7163.2
	$\mathbf{R}^{H}_{M\times M}$	ID	1	7109.1	7148.8
		DIAG	2	7105.5	7150.8
		UNIF	2	7002.5	**7047.9**
		CS_{Het}	-	NA	NA
		AR1	-	NA	NA
		UNST	3	6998.6	7049.6
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	7002.5	**7047.9**
		DIAG	2	7004.5	7055.5
		UNIF	-	NA	NA
		CS_{Het}	-	NA	NA
		UNST	3	7006.5	7063.1
FIB	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	6050.5	**6090.1**
		DIAG	-	NA	NA
		UNST	-	NA	NA
	$\mathbf{R}^{H}_{M\times M}$	ID	1	6050.5	6090.1
		DIAG	2	5997.2	6042.5
		UNIF	2	5985.8	6031.1
		CS_{Het}	-	NA	NA
		AR1	-	NA	NA
		UNST	3	5923.5	**5974.5**
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	5923.5	5974.5
		DIAG	2	5912.1	**5968.7**
		UNIF	2	5922.8	5979.4
		CS_{Het}	-	NA	NA
		UNST	3	5911.2	5973.5
TCH	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	36511.5	36549.2
		DIAG	2	36508.7	36552.7
		UNST	3	36498.7	**36549.0**
	$\mathbf{R}^{H}_{M\times M}$	ID	1	36498.7	36549.0
		DIAG	3	36487.1	36550.0
		UNIF	2	35980.8	**36037.4**
		CS_{Het}	4	35970.1	36039.3
		AR1	2	36090.6	36147.2
		UNST	6	35966.4	36048.2
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	35980.8	36037.4
		DIAG	3	35961.1	**36030.3**
		UNIF	2	35978.9	36041.8
		CS_{Het}	4	35959.0	36034.5
		UNST	6	35960.9	36048.9

Supplemental Table S2 - Continued.

Trait	$\mathbf{R}$ matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{LI}$)	Model	SP80-3280 x RB835486 (SR1) n_{PAR}	AIC	BIC
TPH	$\mathbf{R}_P = \mathbf{R}^{L}_{N\times N} \otimes \mathbf{R}^{H}_{M\times M}$				
	$\mathbf{R}^{L}_{N\times N}$	ID	1	11646.2	**11685.8**
		DIAG	2	11647.1	11692.4
		UNST	3	11641.9	11692.8
	$\mathbf{R}^{H}_{M\times M}$	ID	1	11646.2	11685.8
		DIAG	2	11647.4	11692.7
		UNIF	2	11496.5	**11541.8**
		CS_{Het}	-	NA	NA
		AR1	-	NA	NA
		UNST	3	11496.3	11547.3
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	11496.5	11541.8
		DIAG	2	11490.4	**11541.3**
		UNIF	2	11495.0	11546.0
		CS_{Het}	-	NA	NA
		UNST	3	11488.8	11545.4

Supplemental Table S3 - Selected models for the **R** matrix considering each trait separately. The models for the **R** matrix were selected according to the lowest value of the BIC criterion for the SD, SH, SN, SW, BRIX, POL%C, POL%J, FIB, TCH and TPH for family of sugarcane SR2 at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012 and 2013).

			SP81-3250 x RB925345 (SR2)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_p \otimes \mathbf{R}_K \otimes \mathbf{I}_{I,I}$)	Model	n_{PAR}	AIC	BIC
SD	$\mathbf{R}_p = \mathbf{R}^{E}_{P\times P}$	ID	1	20096.9	20245.3
		DIAG	6	19508.9	**19689.7**
		AR1	5	20096.1	20251.0
		FA1	12	19507.0	19726.5
		UNIF	2	20094.4	20249.3
		CS_{Het}	7	**19504.4**	19691.6
		UNST	21	19514.1	19791.7
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	19508.9	19689.7
		DIAG	3	19372.2	19565.9
		AR1	2	19482.3	19669.5
		FA1	-	NA	NA
		UNIF	2	19490.3	19677.5
		CS_{Het}	4	19358.3	19558.4
		UNST	6	**19335.7**	**19548.8**
SH	$\mathbf{R}_p = \mathbf{R}^{E}_{P\times P}$	ID	1	728.2	876.7
		DIAG	-	NA	NA
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	**702.6**	**857.5**
		CS_{Het}	-	NA	NA
		UNST	-	NA	NA
	$\mathbf{R}_K = \mathbf{R}^{R}_{K\times K}$				
	$\mathbf{R}^{R}_{K\times K}$	ID	1	**702.6**	**857.5**
		DIAG	-	NA	NA
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	704.0	865.4
		CS_{Het}	-	NA	NA
		UNST	-	NA	NA
SN	$\mathbf{R}_p = \mathbf{R}^{E}_{P\times P}$	ID	1	45429.0	45519.6
		DIAG	6	44441.9	44564.7
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	45310.0	45406.9
		CS_{Het}	7	44268.3	44397.6
		UNST	21	**43966.4**	**44186.2**

Supplemental Table S3 - Continued.

			SP81-3250 x RB925345 (SR2)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{LJ}$)	Model	n_{PAR}	AIC	BIC
SN	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	43966.4	44186.2
		DIAG	3	43953.0	44185.8
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	43956.9	44183.1
		CS_{Het}	4	**43943.4**	**44182.6**
		UNST	6	43946.8	44198.9
SW	$\mathbf{R}_P = \mathbf{R}^{\mathbf{L}}_{N\times N} \otimes \mathbf{R}^{\mathbf{H}}_{M\times M}$				
	$\mathbf{R}^{\mathbf{L}}_{N\times N}$	ID	1	47036.9	47101.5
		DIAG	2	**46497.2**	**46568.3**
		UNST	3	46494.7	46572.3
	$\mathbf{R}^{\mathbf{H}}_{M\times M}$	ID	1	46497.2	46568.3
		DIAG	2	46410.5	46494.5
		AR1	5	46107.1	46184.7
		FA1	4	45954.0	46057.4
		UNIF	2	46064.3	46141.8
		CS_{Het}	3	45958.0	46048.5
		UNST	3	**45954.0**	**46057.4**
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	45954.0	46057.4
		DIAG	3	45954.7	45954.7
		UNIF	2	**45928.9**	46038.8
		CS_{Het}	4	45929.6	46052.4
		UNST	6	46065.3	**45929.5**
BRIX	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	7767.2	**7867.1**
		DIAG	6	**7764.0**	7887.4
		AR1	-	NA	NA
		FA1	12	7764.0	7916.7
		UNIF	2	7767.9	7873.7
		CS_{Het}	7	7764.5	7893.8
		UNST	21	7771.5	7953.6
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	**7767.2**	**7867.1**
		DIAG	2	7769.2	7875.0
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	7769.2	7874.9
		CS_{Het}	3	7771.2	7882.8
		UNST	3	7771.2	7882.8

Supplemental Table S3 - Continued.

			SP81-3250 x RB925345 (SR2)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{LI}$)	Model	n_{PAR}	AIC	BIC
POL%C	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	**7653.2**	**7753.0**
		DIAG	5	7637.8	7761.2
		AR1	-	NA	NA
		FA1	10	7641.7	7794.4
		UNIF	2	7654.9	7760.6
		CS_{Het}	6	7639.7	7769.0
		UNST	15	7648.8	7830.9
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	**7653.2**	**7753.0**
		DIAG	2	7651.0	7756.8
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	7655.2	7760.9
		CS_{Het}	3	7653.0	7764.7
		UNST	3	7653.0	7764.7
POL%J	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	8637.5	**8737.4**
		DIAG	5	**8628.1**	8751.5
		AR1	-	NA	NA
		FA1	10	8631.4	8784.1
		UNIF	2	8639.2	8745.0
		CS_{Het}	6	8629.9	8759.2
		UNST	15	8637.6	8819.8
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	**8637.5**	**8737.4**
		DIAG	2	8638.6	8744.4
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	8639.4	8745.1
		CS_{Het}	3	8640.5	8752.1
		UNST	3	8640.5	8752.1
FIB	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	6630.1	6730.0
		DIAG	5	6489.8	**6613.1**
		AR1	-	NA	NA
		FA1	10	6488.5	6641.3
		UNIF	2	6627.7	6733.5
		CS_{Het}	6	**6486.0**	6615.3
		UNST	15	6491.5	6673.6
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	6489.8	**6613.1**
		DIAG	2	**6489.7**	6619.0
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	6490.9	6620.1
		CS_{Het}	3	6490.9	6626.0
		UNST	3	6490.9	6626.0

Supplemental Table S3 - Continued.

			SP81-3250 x RB925345 (SR2)		
Trait	**R** matrix ($\mathbf{R} = \mathbf{R}_P \otimes \mathbf{R}_K \otimes \mathbf{I}_{I,I}$)	Model	n_{PAR}	AIC	BIC
TCH	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	45244.09	45392.66
		DIAG	6	45085.87	45266.74
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	45043.74	45198.77
		CS_{Het}	7	44857.6	45044.92
		UNST	21	**44621.94**	**44899.71**
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	44621.94	44899.71
		DIAG	3	44624.19	44914.87
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	**44605.1**	**44889.32**
		CS_{Het}	4	44607.23	44904.37
		UNST	6	44609.08	44919.13
TPH	$\mathbf{R}_P = \mathbf{R}^{\mathbf{E}}_{P\times P}$	ID	1	15411.74	15546.72
		DIAG	6	15400.97	15565.29
		AR1	-	NA	NA
		FA1	12	15284.83	15484.36
		UNIF	2	15319.87	**15460.72**
		CS_{Het}	7	15311.65	15481.84
		UNST	21	**15262.82**	15515.17
	$\mathbf{R}_K = \mathbf{R}^{\mathbf{R}}_{K\times K}$				
	$\mathbf{R}^{\mathbf{R}}_{K\times K}$	ID	1	**15319.87**	**15460.72**
		DIAG	2	15321.41	15468.13
		AR1	-	NA	NA
		FA1	-	NA	NA
		UNIF	2	15320.62	15467.34
		CS_{Het}	3	15322.22	15474.8
		UNST	3	15322.22	15474.8

Supplemental Table S4 - Average BLUP values to full-sib genotypes with standard deviation (S) and BLUP of the parental genotypes to families SR1 and SR2 and RB867515, the cultivar that was used as a non-parental check, for stalk height (SH) in m, stalk number (SN) by direct counting, stalk diameter (SD) in mm, stalk weight (SW) in kg, cane yield (TCH) in t ha^{-1}, POL%Cane (POL%C), BRIX as °Brix, POL%Juice (POL%J), fiber (FIB) as a percentage and sucrose yield (TPH) in t ha^{-1} at two locations (Araras and Ipaussu, Brazil) over three harvest years (2011, 2012 and 2013).

		Traits									
		SH	SN	SD	SW	TCH	POL%C	BRIX	POL%J	FIB	TPH
	Ave. BLUP	2.31	105.85	28.39	184.32	162.07	15.52	20.85	18.44	12.20	23.72
	S	0.11	15.89	1.71	22.64	12.35	0.61	0.66	0.72	0.63	1.95
SR1	RB867515	2.32	109.80	28.33	185.30	99.50	15.02	20.47	17.83	12.41	15.98
	SP80-3280	2.37	108.10	29.57	203.50	95.84	15.72	21.19	18.63	12.04	16.00
	RB835486	2.22	96.20	27.16	152.30	85.48	15.84	21.05	18.85	12.34	15.00
	Ave. BLUP	2.34	117.44	25.61	168.15	149.04	15.54	20.93	18.57	12.64	22.09
	S	0.12	21.97	1.92	24.81	21.67	0.70	0.79	0.87	0.70	2.61
SR2	RB867515	2.28	109.50	25.41	158.30	141.10	14.82	20.16	17.69	12.64	19.36
	SP81-3250	2.25	127.60	25.25	172.50	150.70	15.60	21.02	18.58	12.37	22.91
	RB925345	2.49	115.20	25.18	175.70	157.00	16.24	21.63	19.47	12.94	24.26

CAPÍTULO 2

GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane

Thiago Willian Almeida Balsalobre[a,b &], Guilherme da Silva Pereira[c &], Gabriel Rodrigues Alves Margarido[c], Rodrigo Gazaffi[a], Fernanda Zatti Barreto[a], Carina de Oliveira Anoni[c], Cláudio Benício Cardoso-Silva[b], Estela Araújo Costa[b], Melina Cristina Mancini[b], Hermann Paulo Hoffmann[a], Anete Pereira de Souza[b], Antonio Augusto Franco Garcia[c], Monalisa Sampaio Carneiro[a]*

[a]Centro de Ciências Agrárias, Departamento de Biotecnologia e Produção Vegetal e Animal, Universidade Federal de São Carlos, Rodovia Anhanguera, Km 174, Araras, CEP 13600-970, São Paulo, Brasil; [b]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Avenida Monteiro Lobato 255, Campinas, CEP 13083-862, SP, Brasil, e Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Avenida Candido Rondon 400, Campinas, CEP 13083-875, SP, Brasil; [c]Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Avenida Pádua Dias 11, Piracicaba, CEP 13418-900, São Paulo, Brasil; [&]These authors contributed equally to this work. *Corresponding author: monalisa@cca.ufscar.br.

Abstract

Background: Sugarcane (*Saccharum* spp.) is predominantly an autopolyploid plant with a variable ploidy level, frequent aneuploidy and a large genome that hampers investigation of its organization. Genetic architecture studies are important for identifying genomic regions associated with traits of interest. However, due to the genetic complexity of sugarcane, the practical applications of genomic tools have been notably delayed in this crop, in contrast to other crops that have already advanced to marker-assisted selection (MAS) and genomic selection. High-throughput next-generation sequencing (NGS) technologies have opened new opportunities for discovering molecular markers, especially single nucleotide polymorphisms (SNPs) and insertion-deletion (indels), at the genome-wide level. The objectives of this study were to (i) establish a pipeline for identifying variants from genotyping-by-sequencing (GBS) data in sugarcane, (ii) construct an integrated genetic map with GBS-based markers plus target region amplification polymorphisms and microsatellites, (iii) detect QTLs related to yield component traits, and (iv) predict putative candidate genes underlying the mapped QTLs.

Results: We used four pseudo-references to align the GBS reads. Depending on the reference, from 3,433 to 15,906 high-quality markers were discovered, and half of them segregated as single-dose markers (SDMs) on average. In addition to 7,049 non-redundant SDMs from GBS, 629 gel-based markers were used in a subsequent linkage analysis. Of 7,678 SDMs, 993 were mapped. These markers were distributed throughout 223 linkage groups, which were clustered in 18 homo(eo)logous groups (HGs), with a cumulative map length of 3,682.04 cM and an average marker density of 3.70 cM. We performed QTL mapping of four traits and found seven QTLs. Our results suggest the presence of a stable QTL across locations. Furthermore, QTLs to soluble solid content (BRIX) and fiber content (FIB) traits had associated markers with candidate genes.

Conclusions: This study is the first to report the use of GBS for large-scale variant discovery and genotyping of a mapping population in sugarcane, providing several insights regarding the use of NGS data in a polyploid, non-model species. The use of GBS generated a large number of functional markers and still enabled ploidy and allelic dosage estimation. Moreover, we were able to identify seven QTLs, two of which had great potential for validation and future use for molecular breeding in sugarcane.

Keywords: *Saccharum* spp., polyploidy, SNPs, molecular markers, allelic dosage, quantitative traits

1. Background

Sugarcane (*Saccharum* spp.) has a complex genome because of its variable ploidy level, frequent aneuploidy and large genome of approximately 10 Gb [1-5]. This crop is a member of the Poaceae family and the Andropogoneae tribe, which includes maize and sorghum [6,7]. Modern sugarcane cultivars are the result of interspecific crosses between the domesticated species *Saccharum officinarum* L. (2n = 80) and the wild species *S. spontaneum* L. (2n = 40–120), followed by several backcrosses with *S. officinarum* [6,8]. These cultivars have chromosome numbers ranging from 100 to 130, are vegetatively propagated, and result from the selection of populations derived from outcrossing heterozygous parents [1,9]. Furthermore, sugarcane has a very high photosynthetic efficiency and is a crop with major economic importance in many tropical and subtropical countries primarily because of its use in the production of sugar and bioethanol [10-12].

Polyploidy, an important driver of plant evolution in natural populations, has played a crucial role in the domestication of crops such as wheat, maize, cotton and potato [13-16]. Sugarcane is predominantly an autopolyploid plant, and the understanding of its genome organization is limited [4,7]. One possible way to increase knowledge of the genome organization of this species is by using genetic maps. High-resolution genetic linkage mapping is essential for quantitative trait loci (QTL) studies in mapping populations and may be a first step toward potential marker-assisted selection (MAS) in plants [17-22]. Several genetic linkage maps of sugarcane have been generated since a methodology based on single-dose markers (SDMs) was proposed by Wu et al. [23]. SDMs that segregate 1:1 and 3:1 in full-sib progenies (F_1 populations) [24] or 3:1 in populations created by selfing an individual are commonly used for constructing genetic maps in sugarcane [4,25-35]. An integrated map of sugarcane with different types of molecular markers, such as microsatellites or single sequence repeats (SSRs) and target region amplification polymorphism (TRAP), extended the characterization of polymorphic variation throughout the entire genome [36-38]. However, in outcrossing heterozygous species such as sugarcane, for each segregating loci, different numbers of segregating alleles can exist, and a relative large number of markers is required to guarantee reasonable coverage of its genome [37,39,40].

Currently, high-throughput next-generation sequencing (NGS) technologies have provided new opportunities for discovering molecular markers, especially single nucleotide polymorphisms (SNPs), at the genome-wide level [41-43]. Some of these techniques, *e.g.*, restriction-site associated DNA sequencing (RAD-seq) [44] and genotyping-by-sequencing

(GBS) [42,45], employ a reduced genome representation that is achieved through restriction enzyme digestion, which could be helpful for a complex genome such as that of a polyploid [46]. Moreover, these strategies require no prior knowledge of the variants being analyzed, making them useful for genetic analysis in species with no reference genome [47]. The GBS protocol has been widely used in a range of genetic studies in several species such as apple, barley, lettuce, switchgrass, maize, rice, wheat, and soybean [42,45,48-54]. SNP datasets generated from GBS can be analyzed to detect associations between genotypes and phenotypes, perform diversity analyses, and construct genetic maps, among other applications [39,55-57].

QTL mapping in sugarcane is a promising tool for characterizing the genetic architecture of several yield component traits of interest, such as sucrose yield, cane yield, stalk diameter, stalk height, stalk number, and stalk weight, as well as resistance to diseases, pests and abiotic stresses [10,58-61]. Sugarcane is a semi-perennial crop with repeated measures data obtained for several harvests and locations, and QTL mapping studies are usually performed in two steps. First, adjusted phenotypic means are obtained; second, these means are searched for associations with molecular markers and/or along genetic maps [31,32,62]. Gazaffi et al. [61] proposed a method that considers an integrated genetic map in which QTL mapping is performed based on the advantages of the composite interval mapping (CIM) approach [63]. Briefly, a mapping model with three genetic effects is considered for genome scanning [61]. It is assumed that a QTL may also segregate in different patterns in progeny as a function of its genetic effects and of the linkage phase between markers and QTL alleles.

This study is the first to report on the development and application of GBS for mapping studies in sugarcane. Our objectives were to (i) establish a pipeline for identifying SNPs and insertion-deletion (indels) from GBS data in a sugarcane F_1 population, (ii) construct the first GBS-based integrated genetic map with additional SSR and TRAP markers in this bi-parental mapping population, (iii) identify QTLs related to yield component traits based on the integrated genetic map, and (iv) predict putative candidate genes underlying mapped QTLs that may be involved in yield traits in sugarcane. We discuss these results in the context of where GBS is likely to be most useful in sugarcane crop development.

2. Material and methods

2.1 Mapping population and DNA extraction

The mapping population consisted of 151 full sibs derived from a commercial cross between the SP80-3280 (female parent) and RB835486 (male parent) sugarcane cultivars. The parents are broadly cultivated throughout Brazil because of their high biomass and sugar yields. SP80-3280 (SP71-1088 × H57-5028) was one of the cultivars with transcriptome sequencing performed previously by SUCEST [64] and RNA-seq [65] projects; its genome is currently being completely sequenced by the Brazilian initiative [66]. This cultivar is resistant to brown rust (*Puccinia melanocephala*), whereas RB835486 (L60-14 × ?) is susceptible to fungal disease. The parents have been used in studies of evolutionary relations in putative tandem gene duplication [67] and retrotransposon-based insertion polymorphisms [68]. Total genomic DNA samples from parents and progeny were extracted from the 1+ internode (leaf primordia) as proposed by Al-Janabi et al. [69], with modifications.

2.2 GBS-based markers

GBS was performed by the Institute for Genomic Diversity (Cornell University, Ithaca, NY, USA) according to the protocol described in detail by Elshire et al. [45]. Samples from both parents of the population were replicated three times for sequencing. Each individual within a library was part of a 96-plex reaction (including one blank sample each). To provide a higher allele depth, libraries were obtained by digestion with *Pst*I, a partially methyl-sensitive six-base-pair site restriction enzyme. Additionally, the 96-plex libraries were run in two distinct lanes each on a HiSeq™ 2000 platform (Illumina® Inc., San Diego, CA, USA).

To discover polymorphisms, we initially used the TASSEL-GBS pipeline [70], which was implemented in TASSEL software (v. 4.3.8). Because this pipeline requires a reference genome and because the complete sugarcane genome sequencing is in progress [66], we proposed the use of four alternative pseudo-references: (i) a methyl-filtered sugarcane genome (~674 Mb arranged in 1,109,444 scaffolds) [71], (ii) the *Sorghum bicolor* genome (v. 2.1; ~726 Mb arranged in 10 chromosomes and 1,600 unassembled scaffolds) [72], (iii) an RNA-seq sugarcane transcriptome (~780 Mb arranged in 119,768 transcripts) [65], and (iv) sequences from the SUCEST project (~152 Mb of a total of 237,954 sequences) [64]. The BOWTIE2 (v. 2.2.1) algorithm was used to map the 64-bp-long tags against each reference. The exact reference and alternative allele depths (read counts) were recorded in variant call

format (VCF) files. To perform this task, we modified the GBS-TASSEL pipeline to record a maximum value of 32,767 counts for each allele.

2.2.1 Allelic dosage estimation and marker curation

The GBS technique generated allele-specific read count data in the form $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ for each biallelic locus from individuals $i = 1,2, \ldots, n$. Data D from each locus were analyzed in SUPERMASSA software [73]. As a prior quality control, markers with more than 25% missing data were filtered out. We also excluded GBS loci data with fewer than 50 read counts for the reference allele on average. In addition, individual data points with the radial coordinate $r_i = \sqrt{x_i^2 + y_i^2}$ smaller than $(0.10) \times max(r_1, r_2, \ldots, r_n)$ were removed.

All even-numbered ploidy levels ranging from 2 to 20 were tested [39]. The ploidy that returned the highest likelihood was selected after fitting a subjacent F_1 segregation model into SUPERMASSA. The replicated parental data provided additional constraints during estimation. Following the recommendation reported in Serang et al. [73] to find the maximum *a posteriori* (MAP) solution for the estimates, the SUPERMASSA naive posterior report threshold was set to zero. Afterward, the values of individual posterior probability given the selected ploidy (6 through 14) were also calculated; these values indicated the maximum threshold that would allow individual assignment to a certain dosage cluster. Only ploidies ranging from 6 to 14 were selected because they were more likely to appear in the sugarcane genome and exhibited a greater number of SDMs [39].

For posterior quality control, we only selected SDMs for which the median of all individual posterior probabilities was higher than 0.80. The SUPERMASSA dosage outputs were recoded for mapping purposes in R software by substituting the respective reference and alternative codominant alleles for *a* and *b*. Redundant loci within each reference and between references were inspected and excluded based on the recoded genotype calls. Here, only non-redundant loci were used for linkage mapping analysis. A circular plot was used to summarize the duplicate loci within and between references using the R CIRCLIZE package [74].

2.3 Gel-based SSR and TRAP markers

A total of 120 SSR markers were genotyped in the 151 full-sib progeny and in the two parents. SSR markers were derived from both ESTs and genomic sequences. There were 98 EST-SSRs named SCA [75,76], SCB [75], SCC [75,77] and IISR [78], and 16 genomic SSRs named SMC [79] and CIR [80]. In addition, there were six SSR markers (named SB, Xtxp, CNL and SvPEPCAA [81-83]) from a genic sorghum library. PCRs were performed in a final volume of 20 μL as described by Oliveira et al. [75].

For TRAP markers four fixed and three arbitrary primers (named ARB1, ARB2 and ARB3) were used. The arbitrary primers were adapted of Li and Quiros [84]. Two fixed primers were designed from *sucrose phosphate synthase* (SuPS) [85,86], and one primer each was designed from *caffeic acid 3-O-methyltransferase* (COMT) and *cinnamoyl-CoA reductase* (CCR) [87] gene sequences. PCRs were performed in a final volume of 20 μL [88].

Amplicons of SSR and TRAP markers were denatured at 90 °C for 3 min in an equal volume of loading buffer (formamide containing 0.8 mM EDTA and traces of bromophenol blue and xylene cyanol), snap-cooled on ice, and electrophoresed in 6% denaturing polyacrylamide gels in 1X TBE buffer. The samples were loaded on a dual vertical electrophoresis system (CBS Scientific) and were run at 75 W for 1 to 3 h depending on the fragment sizes to be separated. A 10-bp ladder was used as a standard size. The bands were visualized by silver staining according to Creste et al. [89].

2.4 Linkage map construction and homo(eo)logous group assignment

Linkage mapping analysis was performed over non-duplicated SDMs using the ONEMAP (v. 2.0-4) R package [90]. This analysis allowed simultaneous estimation of linkage and linkage phases between markers [91] and marker ordering using multipoint likelihood through hidden Markov models [92,93] from a mixed set of different marker segregation patterns. The markers were coded according to the notation proposed by Wu et al. [91]. In brief, the codominant alleles were coded as *a* and *b*, while the null alleles were coded as *o* and treated as recessive alleles. GBS-based codominant markers were used to assess segregation with the following three cross types: ‘B3.7’ (*ab* × *ab*), ‘D1.10’ (*ab* × *aa*) and ‘D2.15’ (*aa* × *ab*). In addition, SSR and TRAP gel-based dominant markers were used to assess three more cross types that are traditionally used in integrated sugarcane maps: ‘C.8’ (*ao* × *ao*), ‘D1.13’ (*ao* × *oo*) and ‘D2.18’ (*oo* × *ao*). ‘D1’ and ‘D2’ stand for crosses in which the marker locus is heterozygous (and hence informative) only to SP80-3280 or to RB835486, respectively; they are both expected to segregate in a 1:1 ratio. ‘B3’ and ‘C’ stand for crosses in which the

marker locus is heterozygous and symmetric in both parents; the former is expected to segregate in a 1:2:1 ratio, whereas the latter will segregate in a 3:1 ratio. Because SUPERMASSA is able to predict parental genotypes using the population data even when parental data are missing, all B3-type markers could be recovered. However, D1- and D2-type markers were only recovered when read counts for at least one parental were available; with no parental data, these markers were discarded.

For gel-based markers, chi-square tests were conducted in R software according to the expected segregation ratios inferred through parental genotypes, and then *p*-values were corrected using false discovery rate control for non-dependent tests as implemented in the 'p.adjusted' R function. For GBS-based markers, segregation had already been considered during dosage estimation in SUPERMASSA software according to the F_1 model [73].

To obtain the genetic map, we first performed a two-point test to identify linkage groups (LGs). Any pairwise markers that showed a ***LOD*** score > 9.0 and a recombination fraction < 0.10 were considered linked. Afterward, we applied ordering algorithms to each group. For the groups with less than six markers, the best order was obtained by performing an exhaustive search with the 'compare' function. For those groups with more than six markers, the 'order.seq' command was used, *i.e.*, an initial set of the five most informative markers (preferentially B3- and C-type markers) was sampled for an exhaustive search. The best order was used as a frame for the consecutive inclusion of new markers. Once these groups were obtained, we used the 'try.seq' function to verify markers that were considered unlinked according to the initial procedure, and it was possible to integrate the pre-ordered groups. In this step, the following other markers were also tested: (i) markers at the ends of the LGs more than 20 centiMorgans (cM) far from the closest marker; and (ii) markers belonging to very small LGs (with sizes less than 1 cM or containing only two loci). As a final step, the LGs with more than five markers were refined using the 'ripple' algorithm within a sliding window of five markers. The ordered group heatmap plots were inspected visually, and manual correction was performed when needed throughout the map building process. The LGs were drawn in MAPCHART software [94]. The homo(eo)logous groups (HGs) were defined according to the sugarcane reference scaffolds shared by the GBS-based markers. Gel-based markers were also checked because they can produce different alleles that share the same primer pair.

2.5 Phenotypic data

The mapping population was planted in 2010 at two locations (Araras, located at 22°21'25" S, 47°23'03" W, and Ipaussu, located at 23°08'44" S, 49°23'23" W; both in the State of Sao Paulo, Brazil) and evaluated during three harvest years for several yield component traits, including sucrose content of cane (POL%C, in %), soluble solid content (BRIX, in °Brix), stalk diameter (SD, in mm) and fiber (FIB, in %). At each location, the experimental design consisted of an augmented randomized incomplete block design, which was fully replicated three times. For each trait, a multiple-harvest-location trial was considered under a mixed linear model approach for each yield component [10].

2.6 QTL mapping

The joint adjusted phenotypic means by location for each trait were used for QTL mapping. The QTL mapping methodology applied in this work was presented by Gazaffi et al. [61], and it expands the CIM method [63] to full-sib families. In brief, the model has three genetic effects, with two for additive effects (one for each parent) and one dominance effect (intra-loci interaction). To infer the conditional probabilities of QTL genotypes, multipoint probabilities were obtained using hidden Markov models at each 1 cM from the genetic map.

The mapping strategy was based on three steps. First, an interval mapping (IM) [95] search was carried out in order to select marker cofactors. The peaks with a LOD score greater than 2 were sampled for inclusion in the QTL detection procedure. If the peak was not coincident with a marker, the closest one was considered as cofactor. Second, the QTL search was performed along the genome and considered the cofactors located outside the linkage group under analysis. To declare a QTL, the threshold for each search was obtained from 1,000 permutations with a significance level of 0.95 [96]. Finally, the peaks above the permutation threshold were fully characterized, *i.e.*, the significance of each genetic effect was tested along with the linkage phase between markers and QTLs and the QTL segregation pattern. The proportions of phenotypic variance (R^2) as explained by each detected QTL were obtained for all the effects simultaneously. All the analyses were performed in R software [97].

2.7 Candidate gene identification

Functional annotation of the regions of adjacent markers from the mapped QTLs was performed using sequence information from the scaffolds of the methyl-filtered sugarcane genome, sugarcane transcriptome from RNA-seq assembly, sequences from the SUCEST project and sequences with 400 nucleotides in length at both sides of the SNP/indel position for mapped markers from the sorghum genome. These scaffolds and sequences were annotated using Blast2GO software version 3.1 [98] on the non-redundant NCBI database with E values $\leq 1 \times 10^{-3}$, and the Phytozome website [99] was used to align the data against the *Viridiplantae* protein databases.

3. Results

3.1 GBS-based marker polymorphism discovery

Short-read sequences were obtained from the mapping population and triple-replicated parents after double sequencing the 96-plex *Pst*I libraries. Of the 330 million good barcoded reads, more than 3.1 million resulting tags were obtained for alignment against four different pseudo-references. Three of four pseudo-references originated from sugarcane DNA or RNA libraries. The methyl-filtered sugarcane genome resulted in the highest alignment rate, with 87.94% (2,729,457) aligned tags. Regarding RNA-based references, 38.53% (1,195,723) and 23.89% (741,537) tags were aligned with the RNA-seq transcriptome and SUCEST project sequences, respectively; however, their rates of non-unique alignment differed greatly (Supplementary Table 1). Finally, the reference for the close-relative genome of sorghum had 42.29% (1,312,661) aligned tags.

From 39,058 to 151,755 biallelic variants were identified, depending on the reference. Furthermore, for all the references, SNPs were identified more often than indels, at a 2.6:1 ratio on average (Table 1). With respect to SNPs, transitions (purine-purine or pyrimidine-pyrimidine interchanges) were identified 1.4 times more often than transversions (purine-pyrimidine interchanges). Approximately 12% of the markers with more than 25% missing data were filtered out per reference (Table 1) because current missing data imputation methods are not able to handle the complexity of the sugarcane genome. Low-coverage or ambiguous loci were also broadly present. At this stage of analysis, ~64% underrepresented loci were also excluded by considering a minimum of 50 read counts on average for the reference alleles. Although a large number of loci were removed, from 8,885 to 38,378 high-coverage loci, missing-filtered polymorphic loci were subjected to SUPERMASSA quantitative genotyping analyses. The remaining redundancy was investigated only after these analyses.

3.2 Ploidy and allelic dosage estimation

Ploidy levels ranging from 2 to 20 were evaluated with SUPERMASSA software. Once the more likely ploidy was acknowledged, the software provided the individual posterior probability for each individual that was allocated in one of the expected dosage clusters. For the ploidy levels considered in these analyses, the number of loci varied within each ploidy class (Figure 1), and an average of 10.7% loci were classified as having ploidies of 2 or 4, 60.3% as ploidies 6 through 14, and 29.0% as ploidies 16 through 20. Here, we used the same *ad hoc* criteria to classify each locus into one quality category based on their posterior probabilities for each ploidy; categories A and B included loci with either the highest or the sum of the two highest posterior probabilities that were greater than or equal to 0.80, respectively, and category C included all other cases [39]. Categories A, B and C represented 60.4%, 26.6% and 12.9% of the loci on average for all the ploidies, respectively (Figure 1).

Table 1 - Number of markers generated after GBS-TASSEL pipeline analyses to map the GBS sugarcane population data.

GBS-TASSEL pipeline pseudo-references	SNPs	Indels	Total	Excluded data		Filtered polymorphic sites
				Missing data[a]	Low coverage loci[b]	
Methyl-filtered sugarcane genome	110,261	41,494	151,755	16,815 (11.1%)	96,562 (63.6%)	38,378 (25.3%)
Sorghum bicolor genome (v. 2.1)	84,757	35,447	120,204	13,773 (11.5%)	78,914 (65.6%)	27,517 (22.9%)
RNA-seq sugarcane transcriptome	73,275	26,778	100,053	11,809 (11.8%)	63,658 (63.6%)	24,586 (24.6%)
SUCEST project sequences	29,238	9,820	39,058	4,878 (12.5%)	25,295 (64.8%)	8,885 (22.7%)

Notes:
[a] More than 25% of the population is missing data.
[b] Less than 50 reads on average for the reference alleles.

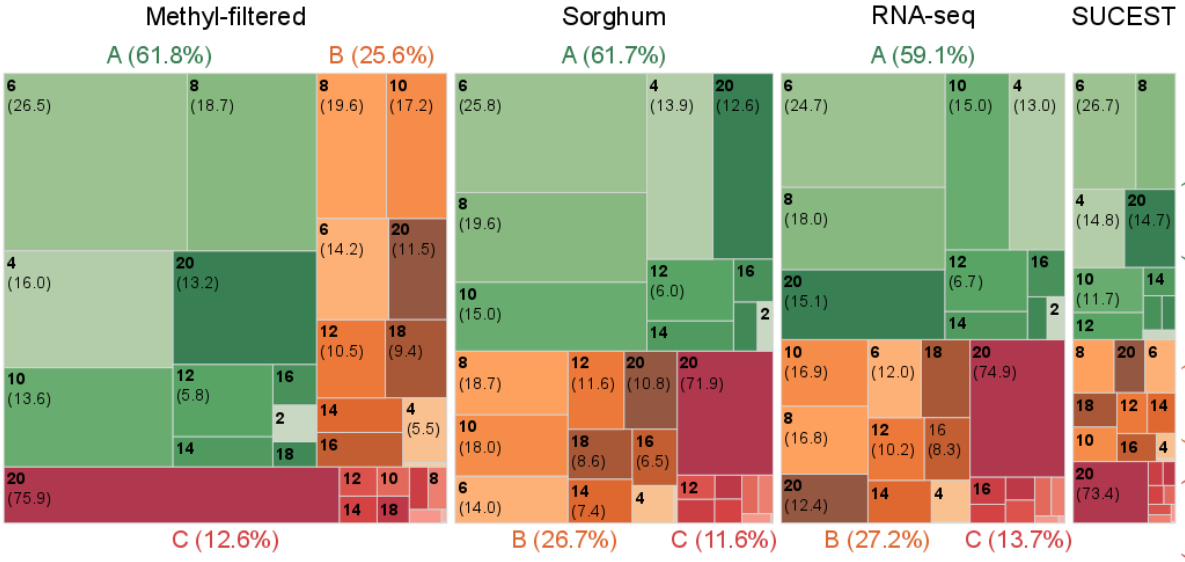


Figure 1 - Mosaic plot showing the ploidy levels that produced the highest posterior probabilities for the mapping of GBS sugarcane population data considering the following four pseudo-references: the methyl-filtered sugarcane genome, *Sorghum bicolor* genome, RNA-seq sugarcane transcriptome and sequences from the SUCEST project. The areas of the rectangles are proportional to the number of loci that have the same ploidy level, as indicated within each rectangle in parentheses. According to the posterior probabilities calculated for each even-numbered ploidy level within a range from 2 to 20, each locus was classified into one category using the following *ad hoc* criteria: Category A (green), when the highest posterior probability was greater than or equal to 0.80; Category B (yellow), when no single value of the posterior probability was higher than 0.80 but the sum of the two highest ones was greater than or equal to 0.80; and Category C (red), which included all other cases. In parentheses: the number of loci as a percentage within the given ploidy level and category.

For the linkage map construction, we selected the loci that were classified into category A and the ploidies ranging from 6 to 14, which represented 40.7% of loci on average from the total input in SUPERMASSA or from 3,433 to 15,906 markers depending on the reference (Table 2). In addition, we characterized these remaining "good-quality loci" according to dosage. SDMs and multi-dose markers (MDMs) were equally represented by GBS, with approximately 50% on average for each one. As a final quality control analysis of the GBS data, we selected the loci with the median of all individual posterior probabilities greater than 0.80. This *ad hoc* criterion aimed to ensure that the loci used in genetic mapping had at least 50% of the individuals with a high probability (superior to 0.80) of being in their given clusters for the selected ploidy. Depending on the number of clusters, the SDMs were classified as segregating in a 1:2:1 ratio (three clusters) or in a 1:1 ratio (two clusters). On average, the percentages that represented each segregation class were 16.4% and 83.6%, respectively (Table 2).

Table 2 - Selected loci with good quality (category A) and ploidy (6 through 14) were classified as single-dose markers (SDMs) or multi-dose markers (MDMs). High-quality SDMs (median of all individual *a posteriori* probabilities > 0.80) were also characterized according to their segregation pattern in the sugarcane mapping population.

Reference	Total	Dosage		High-quality SDM	Segregation pattern	
		MDM	SDM		1:2:1	1:1
Methyl-filtered sugarcane genome	15,906	7,014 (44.1%)	8,892 (55.9%)	5,266	912 (17.3%)	4,354 (82.7%)
Sorghum bicolor genome (v. 2.1)	11,789	5,784 (49.1%)	6,005 (50.9%)	3,433	605 (17.6%)	2,828 (82.4%)
RNA-seq sugarcane transcriptome	9,808	4,959 (50.6%)	4,849 (49.4%)	2,869	469 (16.4%)	2,400 (83.6%)
SUCEST project sequences	3,433	1,736 (50.6%)	1,697 (49.4%)	983	141 (14.3%)	842 (85.7%)

The redundancy of the SDM was inspected after quality and ploidy filtration with the alleles recoded as *a* or *b*. All the references showed very similar levels of redundancy within and between them (Figure 2). For instance, the same overall level of 22.7% for within-redundancy was found. Only 84 SNPs markers were attributed equally to all four references and represented approximately 3.8% of each reference. Interestingly, each reference provided 39.6% new loci on average. By keeping only one call for each ambiguous marker, we obtained 7,049 loci in total for mapping. Of these loci, 5,757 (81.67%) and 1,292 (18.33%) segregated 1:1 and 1:2:1, respectively (Table 3).

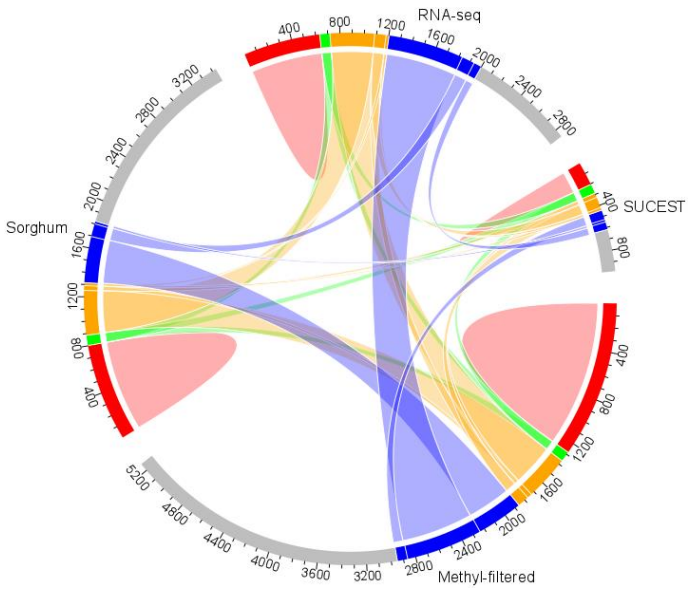


Figure 2 - Circular plot showing the redundancy between single-dose markers from four pseudo-references (methyl-filtered sugarcane genome, *Sorghum bicolor* genome, RNA-seq sugarcane transcriptome and SUCEST project sequences) that were used to align the GBS sugarcane tags. The red regions represent redundancy within each pseudo-reference, whereas

the green, orange and blue regions represent redundancy between four, three and two pseudo-references, respectively. The remaining grey regions represent loci that are unique to each pseudo-reference.

3.3 Gel-based marker genotyping

A total of 120 SSRs and four combinations of TRAP markers (COMT + ARB2, SuPS + ARB1, CCR + ARB1, and CCR + ARB3) produced 1,031 polymorphic bands. Of these 1,031 bands, 545 (52.86%) were tested for 1:1 segregation, and 486 (47.14%) were tested for 3:1 segregation. The number of SDMs feasible for linkage analysis was 629 (61%), of which 506, 84 and 39 originated from genic SSR, genomic SSR and TRAP markers, respectively (Table 3).

3.4 Genetic map

An integrated genetic map was constructed using 151 full sibs generated from a cross between SP80-3280 and RB835486 (Supplementary Figure 1). Of the 8,080 SDMs that were scored (Table 3), 7,678 were used for linkage analysis (7,049 GBS-based markers and 629 gel-based markers), and 993 (12.93%) were placed in the linkage map (Table 3 and Table 4). The mapped markers included 934 GBS-based markers and 59 SSRs. The distribution of the segregation patterns of mapped markers were 254 B3-type markers (1:2:1), 15 C-type markers (3:1), 518 D1-type markers or that were informative only for SP80-3280 (500 GBS-based markers and 18 gel-based markers) and 206 D2-type markers or that were informative only for RB835486 (180 GBS-based markers and 26 gel-based markers) (Table 4). The markers were distributed throughout the 223 LGs, with a cumulative map length of 3,682.04 cM and an average marker density of 3.70 cM (Table 5). The length of LGs ranged from 1.06 cM (LG 70) to 235.67 cM (LG46), with an average of 16.51 cM; 56 LGs displayed lengths shorter than 2 cM, 95 LGs exhibited lengths greater or equal to 2 cM and smaller than 10 cM, and the other 72 LGs had lengths greater or equal to 10 cM.

A total of 18 HGs were formed based on the common genomic origins of mapped loci from different LGs, which were provided by SSR and GBS-based markers. The number of LGs allocated into HGs ranged from two (HG1, HG2, HG4, HG6, HG7, HG8, HG9, HG13, HG14 and HG18) to five (HG11). The coverage within the HGs varied from 2.91 cM (HG13) to 273.66 cM (HG11). A total of 175 LGs with 730 markers remained unassigned to any HG (Table 5 and Supplementary Figure 1).

Table 3 - Overall single-dose gel-based and GBS-based markers screened for the progeny of the cross between sugarcane cultivars SP80-3280 and RB835486.

Markers	Gel-based markers			GBS-based markers	Total
	Genomic SSR	Genic SSR	TRAP		
Number of SDMs evaluated (gel-based and GBS-based markers)	109	842	80	7,049	8,080
SDMs with 1:1 segregation	66	456	23	5,757	6,302
SDMs with 1:2:1 segregation (GBS-based markers)	-	-	-	1,292	1,292
Double SDMs (gel-based markers) with 3:1 segregation	43	386	57	-	486
Number of markers with distorted segregation	25	336	41	0	402
Total number (1:1, 1:2:1 and 3:1) feasible for linkage analysis	84	506	39	7,049	7,678

Table 4 - Distribution of the different marker types as mapped according to their cross type.

Cross type	Number of markers				
	Gel-based markers			GBS-based markers	Total
	Genomic SSR	Genic SSR	TRAP		
D1.10 (*ab x aa*)	-	-	-	500	500
D1.13 (*ao x oo*)	4	14	0	-	18
D2.15 (*aa x ab*)	-	-	-	180	180
D2.18 (*oo x ao*)	4	22	0	-	26
B3.7 (*ab x ab*)	-	-	-	254	254
C.8 (*ao x ao*)	2	13	0	-	15
Total	10	49	0	993	993

Table 5 - Number of each type of mapped marker within each homo(eo)logous group (HG), number of linkage groups (LGs) within each HG, the length of each HG in centimorgans (cM) and the marker density in cM of each HG for the genetic map construct from a progeny of a cross between sugarcane cultivars SP80-3280 and RB835486.

HG	No. LGs	No. SSR	No. GBS-based markers	No. mapped markers	Length of HG (cM)	Marker density (cM)
1	2	0	11	11	48.90	4.44
2	2	0	8	8	61.72	7.71
3	3	5	21	26	157.14	6.04
4	2	0	9	9	59.09	6.56
5	4	5	17	22	100.45	4.56
6	2	2	14	16	144.24	9.01
7	2	1	9	10	21.45	2.14
8	2	0	8	8	37.84	4.73
9	2	4	5	9	53.72	5.96
10	3	0	19	19	120.08	6.32
11	5	13	31	44	273.66	6.22
12	3	5	7	13	40.64	3.12
13	2	0	7	7	2.91	0.41
14	2	0	9	9	60.21	6.69
15	3	0	13	13	69.55	5.35
16	4	10	8	18	120.73	6.70
17	3	0	14	14	15.35	1.09
18	2	4	3	7	37.26	5.32
Unassigned in HG	175	10	721	730	2,257.10	3.09
Total	223	59	934	993	3,682.04	3.70

3.5 QTL mapping

QTL mapping was performed for POL%C, BRIX, SD and FIB traits [10] by applying a CIM model [61] to the integrated genetic map. Considering all traits, 24 cofactors were found for each location, Araras and Ipaussu. The trait with more cofactors was SD, with eight cofactors identified for each location (Supplementary Table 2).

To declare the significant QTLs, a permutation test was performed for each phenotype [96]. The values of the LOD score threshold for BRIX, POL%C, SD and FIB at Araras and Ipaussu were 3.79 and 3.77, 3.80 and 3.86, 4.20 and 4.28, and 4.45 and 4.14, respectively. In this case, we were able to declare seven QTLs. For the respective locations, Araras and Ipaussu, BRIX had two and one QTLs, POL%C had one and one QTL, SD had zero and one QTL, and FIB had one and zero QTLs.

The global LOD score values ranged from 4.17 to 6.02, and the R^2 values ranged from 2.71% to 9.19%. The highest LOD score for the additive effect was 5.22 for parental

SP80-3280 for a QTL associated with BRIX at Araras (B1 at LG4). The highest LOD score for the dominance effect was 3.93 for a QTL associated with FIB at Araras (FIB1 at LG46). The segregation patterns of the QTLs were as follows: 1:1 (42.85%), 1:2:1 (14.30%) and 3:1 (42.85%). The results of the QTL mapping are summarized in Table 6 and Figure 3.

For BRIX, two QTLs (B1 and B2) explained 10.54% of the phenotypic variation in Araras. The QTL identified in Ipaussu (B3 at LG4) was also part of a set found in Araras, *i.e.*, it could be near in the LG at both locations and showed similar effects; this QTL had a significant additive effect for parental SP80-3280 and a segregation pattern of 1:1. In addition, the mapping analysis showed that the region of QTLs B1 and B3 at LG4 was also associated with POL%C (P1 and P2). QTLs for SD and FIB (SD1 and FIB1) showed larger dominance effects that were negative for SD and positive for FIB (Table 6 and Figure 3).

3.6 Candidate gene identification

Sequence homology was found for six out of seven adjacent markers of the mapped QTLs, with homologies for *S. bicolor*, *Solanum tuberosum* and *Zea mays*. A functional description of the sequences showed possible candidate genes for BRIX, SD and FIB traits, whereas the sequence from the marker associated with the QTL found for POL%C did not show homology or a characterized protein. Of the total mapped QTLs, three presented adjacent markers and were located in LG47 (B2), LG29 (SD1) and LG46 (FIB1) for BRIX, SD and FIB traits, respectively (Table 6 and Table 7).

For BRIX, the QTL B2 had two adjacent markers that were each identified in a different reference. The region of the GBS-based marker SCSFAM1074E10_287, which originated from SUCEST sequences, showed homology with *extended synaptotagmin-1-like*, which is a member of a family of membrane-trafficking proteins. The second adjacent GBS-based marker of this QTL, mf16592_3766, which originated from the methyl-filtered sugarcane genome, showed homology with a hypothetical protein in *S. bicolor* (Table 6 and Table 7).

For SD, the QTL SD1 had two adjacent GBS-based markers, which were both identified from the sorghum genome. The markers sb2_61882838 and sb2_61882853 have a small physical distance in the sorghum genome and share almost the same sequence, which

Table 6 - QTLs mapped for BRIX, POL%C, SD and FIB traits by applying a CIM model in Araras (Location 1) and Ipaussu (Location 2).

Location	QTL	Trait	LG[(1)]	Position (cM)[(1)]	Flanking Markers[(2)]	Global LOD[(2)]	R^2 [(3)]	Additive effect SP80-3280[(4)]	LOD[(5)]	Additive effect RB835486[(4)]	LOD[(5)]	Dominance effect[(4)]	LOD[(5)]	Segreg.[(6)]
1	B1	BRIX	4	43.32	mf60753_2013 (QTL)	5.75	9.19	-0.31	5.22	0.08	0.25	0.04	0.09	1:1
1	B2	BRIX	47	40.00	SCSFAM1074E10_287 - QTL - mf16592_3766	4.79	4.78	-0.23	2.22	0.28	3.77	-0.17	1.26	3:1
2	B3	BRIX	4	43.32	mf60753_2013 (QTL)	4.24	7.65	-0.23	3.84	0.06	0.16	0.01	0.01	1:1
1	P1	POL%C	4	43.32	mf60753_2013 (QTL)	4.18	8.10	-0.23	4.14	0.14	0.84	0.08	0.27	1:2:1
2	P2	POL%C	4	43.32	mf60753_2013 (QTL)	4.17	8.09	-0.23	4.13	0.14	0.83	0.08	0.27	1:1
2	SD1	SD	29	2.00	sb2_61882838 - QTL - sb2_61882853	6.02	5.38	0.36	1.66	0.64	4.57	-0.54	3.32	3:1
1	FIB1	FIB	46	171.00	sb1_29954417 - QTL - mf125302_409	4.77	2.71	-0.47	3.11	-0.75	3.95	0.38	3.93	3:1

(1) LG: Linkage groups; Position (cM): QTL position on LG; (2) Adjacent markers for QTLs and associated LODs; (3) Explained phenotypic variation; (4) Additive effects of parents and dominance effects; (5) LODs of the additive and dominance effects; (6) Estimation of the segregation pattern of the QTLs.

Table 7 - Functional description of the sequences that gave rise to adjacent markers of the mapped QTLs for the traits BRIX, POL%C, SD and FIB, and references regarding their functions in plants.

Marker	QTLs	LG	Locations	Traits	Description	e-value	Reference
mf60753_2013	B1, B3, P1, P2	4	1 and 2	BRIX and POL%C	No homology found	-	-
SCSFAM1074E10_287	B2	47	1	BRIX	Extended synaptotagmin-1-like [*Zea mays*]	2.6^{e-26}	[138-140]
mf16592_3766	B2	47	1	BRIX	Hypothetical protein SORBIDRAFT_03 g038130 [*Sorghum bicolor*]	1.2^{e-18}	Unknown function
sb2_61882838	SD1	29	2	SD	Hypothetical protein SORBIDRAFT_02 g026690 [*Sorghum bicolor*]	4.0^{e-13}	Unknown function
sb2_61882853	SD1	29	2	SD	Hypothetical protein SORBIDRAFT_02 g026690 [*Sorghum bicolor*]	4.0^{e-13}	Unknown function
sb1_29954417	FIB1	46	1	FIB	Transposon mutator sub-class [*Sorghum bicolor*]	2.0^{e-177}	[141,142].
mf125302_409	FIB1	46	1	FIB	Zinc finger protein CONSTANS-LIKE 15 [*Solanum tuberosum*]	0.0	[146,149,150]

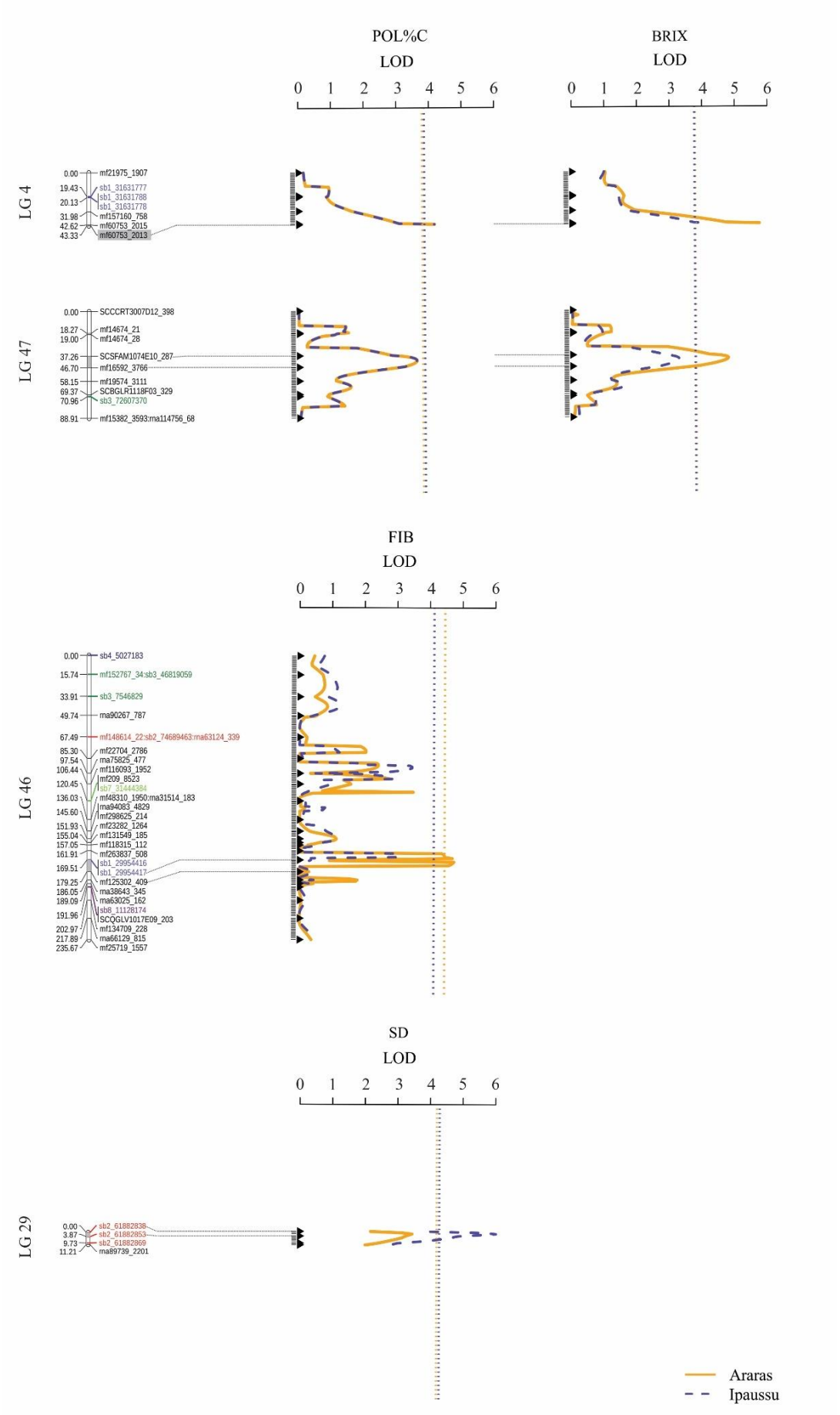

POL%C
LOD
BRIX
LOD
0 1 2 3 4 5 6
LG 4
0.00 mf21975_1907
19.43 sb1_31631777
20.13 sb1_31631788
sb1_31631778
31.98 mf157160_758
42.62 mf60753_2015
43.33 mf60753_2013
LG 47
0.00 SCCCRT3007D12_398
18.27 mf14674_21
19.00 mf14674_28
37.26 SCSFAM1074E10_287
46.70 mf16592_3766
58.15 mf19574_3111
69.37 SCBGLR1118F03_329
70.96 sb3_72607370
88.91 mf15382_3593:rna114756_68
FIB
LOD
0 1 2 3 4 5 6
LG 46
0.00 sb4_5027183
15.74 mf152767_34:sb3_46819059
33.91 sb3_7546829
49.74 rna90267_787
67.49 mf148614_22:sb2_74689463:rna63124_339
85.30 mf22704_2786
97.54 rna75825_477
106.44 mf116093_1952
120.45 mf209_8523
sb7_31444384
136.03 mf48310_1950:rna31514_183
145.60 rna94083_4829
mf298625_214
151.93 mf23282_1264
155.04 mf131549_185
157.05 mf118315_112
161.91 mf263837_508
169.51 sb1_29954416
sb1_29954417
179.25 mf125302_409
186.05 rna38643_345
189.09 rna63025_162
sb8_1128174
191.96 SCQGLV1017E09_203
202.97 mf134709_228
217.89 rna66129_815
235.67 mf25719_1557
SD
LOD
0 1 2 3 4 5 6
LG 29
0.00 sb2_61882838
3.87 sb2_61882853
9.73 sb2_61882869
11.21 rna89739_2201
Araras
Ipaussu

Figure 3 - Composite interval mapping (CIM) for soluble solid content (BRIX, in °Brix), sucrose content of cane (POL%C, in %), stalk diameter (SD, in mm) and fiber content (FIB, in %) from the SP80-3280 and RB835486 F1 population. Blue and yellow dotted lines indicate the LOD thresholds for Ipaussu-SP and Araras-SP, respectively, obtained after permutation tests. The portions highlighted in gray in the linkage groups show the positions of the QTLs.

was determine by homology analysis. These two markers showed homology with a hypothetical protein in *S. bicolor* (Table 6 and Table 7).

For FIB, the QTL FIB1 had two adjacent GBS-based markers from two distinct references. Moreover, each of the two adjacent markers showed a different homology. The region of the sb1_29954417, which originated from sequences of the sorghum genome, had homology with *transposon mutator sub-class*, and the second adjacent marker, mf125302_409, which originated from the methyl-filtered sugarcane genome, had homology with *zinc finger protein CONSTANS-LIKE 15* (Table 6 and Table 7).

4. Discussion

The simultaneous identification and genotyping of SNPs and indels is possible because of important recent advances in sequencing [41-49]. GBS is the preferred high-throughput genotyping method for plants with some level of genetic complexity; this method involves complexity reduction and multiplex sequencing to produce high-quality polymorphism data at a relatively low cost per sample [100]. Using four pseudo-references to discover GBS-based markers, we obtained more markers suitable for linkage analysis (Table 3) than any other previously published study on sugarcane mapping. The strategies adopted for the discovery of GBS-based markers allowed us to relate the sugarcane markers to sorghum chromosomes [72] and to potential genetic regions sampled from the methyl-filtered sugarcane genome [71], RNA-seq sugarcane transcriptome [65] and SUCEST project sequences [64].

The highest alignment (87.94%) of the 3.1 million resulting tags against the methyl-filtered sugarcane genome also reflects most of the high-quality SDMs (Table 2). This great alignment rate may be related to the greater amount of scaffolds for this reference compared with the other three references and to the fact that the *Pst*I enzyme used for library formation is sensitive to DNA methylation [101]; thus, more polymorphic sites are expected from the methyl-filtered genome. Additionally, GBS-based markers had more markers mapped to parent SP80-3280 than to parent RB835486 (Table 4). This result can be explained

by the possible presence of more *Pst*I enzyme restriction sites in cultivar SP80-3280 than in RB835486, leading to more polymorphic sites in the first cultivar, as shown for barley cultivars by Liu et al. [53]. Furthermore, inconsistencies in the number of sites sequenced per sample [101] and in the number of reads per site [102,103], in addition to the filtering steps applied to the GBS libraries to obtain the markers, can influence the observed result. Other factors that can influence these results are the quality and quantity of the biological replicates used for GBS-based marker calling. Because the sequencing of the samples can present failures that will be included in the downstream process, better dosage and ploidy level estimates for each marker in the SUPERMASSA software can be hampered.

The analysis of the loci with high coverage that were filtered for missing data after analysis using SUPERMASSA software showed that for the ploidy levels under consideration, the number of loci varied within each ploidy class (Figure 1), suggesting that the number of chromosomes within the HGs is not constant in sugarcane, as reported previously [39,104]. As stated by Garcia et al. [39], technique artifacts yielding either strong bias or too much noise should explain marker misclassification, *i.e.*, loci not included in the 6-14 expected ploidy range. In fact, the graphical Bayesian model used in the analyses benefits smaller ploidies due to parsimony when the skewed clusters are confounded or, conversely, favors a higher number of clusters by attempting to explain a diffuse scatterplot [105]. In addition, Garcia et al. [39] hypothesized that poor-quality data can also be generated by biological events such as copy number variations or paralogous regions. We used the same *ad hoc* criteria as Garcia et al. [39] to classify each locus into one quality category based on the *a posteriori* probabilities for each ploidy category A as represented by 60.4% of loci of all ploidies on average (Figure 1), which is smaller than the 77.6% of Sequenom-based data that were previously studied [39]. The GBS read count data worked slightly poorer because of their broad genome coverage and eventual technique artifacts. Despite this reduction, a large part of the loci could be further exploited for mapping purposes.

The presence of repeat elements, paralogs, and incomplete or inaccurate reference genome sequences can create ambiguities in GBS-based marker calling [106]. After we selected the loci that were classified into category A and ploidies ranging from 6 to 14 (Table 2), we continued on to the final steps of quality control and redundancy analyses that showed a low redundancy considering simultaneously all four references. Aitken et al. [35] presented the first sugarcane genetic map with DArT markers and did not remove any redundant markers. Sugarcane has a large and complex genome, and a low level of redundancy is important for showing the true coverage of the genome. Heslot et al. [51] showed that DArT

markers were significantly more redundant than GBS markers, and they suggested that GBS markers were significantly more evenly distributed across the wheat genome. These authors also concluded that GBS is the platform for further genomic selection in addition to diversity analyses.

The integrated genetic map of sugarcane obtained in this paper presents improvements in comparison with previous works. Here, the number of markers used for linkage analysis is more than twice the number of markers used for the development of the largest map [35]. Furthermore, this genetic map of sugarcane is the first to use a high-throughput approach for genotyping. Co-dominant biallelic markers can segregate in a 1:2:1 fashion ('B3' cross type), which is even more informative for map integration purposes. The previously published genetic maps had molecular markers even when the potentially co-dominant were treated as dominant. The mapping population was formed by a cross between polyploid heterozygous parents, and for each segregating loci, there could be different numbers of segregating alleles and different dosages that are potentially expressed. Thus, accessing the dosage information of the SDMs with a segregation pattern of 1:2:1 was important; the double SDMs with a segregation pattern of 3:1 ('C8' type) were also important. This information was used to construct an integrated genetic map for sugarcane that increased the genome coverage.

The 223 LGs obtained here had a cumulative map length of 3,682.04 cM and an average marker density of 3.70. The number of LGs exceeds the number of chromosomes of modern sugarcane cultivars, which can range from 100 to 130 [1,9], and 56 LGs showed lengths shorter than 2 cM. This result indicates that gaps remain in our knowledge of most chromosomes, showing that the map is not well saturated. In 2007, Oliveira and collaborators [38] claimed that because there is a constraint to discarding markers in multiples doses, *i.e.*, duplexes of monoparental origin, triplex or higher multiplex markers, gaps are evidently expected; the same discussion should be applied to this study. The number of unlinked markers (87.07%) is higher than that obtained in other sugarcane maps [4,31,35,37,38,80,107-109] and reflects the highly stringent criteria used to construct an integrated genetic map of sugarcane that is reliable for performing QTL mapping analysis.

To increase the understanding of the genetic architecture of sugarcane, a necessary requirement is the availability of good genetics maps, *i.e.*, maps with a high density of markers and with high coverage of the genome [110-112]. The complexity of the sugarcane genome, the cost of generating a large number of markers, and the absence of a statistical genetic model that could consider other segregation ratios beyond 1:1, 1:2:1 and 3:1 have

limited the development of high-density genetic maps. These limitations have delayed practical applications of genomic tools in sugarcane, in contrast to other crops that have already advanced to MAS and genomic selection. Sugarcane still does not have its genome completely sequenced, and sorghum genome is widely recognized as a reference genome for comparative analysis with sugarcane [67]. The origin of modern sugarcane cultivars raises issues that are related to not only the extent and nature of the divergence of the sugarcane and sorghum genomes but also the relations (in terms of meiosis and dosage) among homo(eo)logous loci [67]. Differences in chromosome structures between the ancestor species and pairing behavior in modern cultivars suggest that the hybrid monoploid number is likely to be greater than 10 in sugarcane hybrids [38,113]. Probably because of aneuploidy, an unequal number of chromosomes in each HG is likely to occur; this inequality was reflected in the 18 HGs with differences in genome coverage. Moreover, translocation events may have occurred in sugarcane between regions equivalent to sorghum, as discussed previously [27,28,114-117]. Although these comparative studies proposed hypotheses about the evolutionary aspects of sugarcane and sorghum, the results showed variations that were primarily derived from the low resolution of the genetic maps used and to the coverage of the sugarcane genome. In addition, it is important to highlight that advances in the assembly of polyploid genomes will enable the use of the full sugarcane genome as a reference in the future [118].

The genetic maps and field data obtained through designed experiments are required for mapping studies. For sugarcane, multiple harvest-location trials may be used to infer the genetic architecture of quantitative traits. However, this inference makes the data analysis more complex and challenging because of the interactions that it generates, *e.g.*, genotype by environment interactions. To solve this problem, a mixed model approach has been used to obtain highly accurate genetic estimates [10,31,119], and for segregating populations, these results will be the input for QTL mapping.

In this study, QTL mapping was performed by applying the statistical model proposed by Gazaffi et al. [61], which extends the CIM [63] for a full-sib progeny. The primary advantage of CIM is that it is more precise and effective at mapping QTLs in comparison with single-marker analysis and IM, especially when QTLs are present outside the mapping window [63]. The results obtained from the CIM method are usually comparable to those obtained from multi-QTL analysis if a high-density genetic map is employed to better represent the number of loci underlying the quantitative traits [78]. Several QTL mapping studies in sugarcane have been published [31,34,60,80,120-135]. The comparison between the

results may be biased by several issues, such as the different rates of polymorphisms in parents, the number of progeny, the evaluation methodologies for phenotypic traits, the methodologies used for QTL detection, genetic map contruction, and experimental design, among others. For example, Pastina et al. [31] worked with a population of 100 individuals from a cross between cultivars SP80-180 and SP80-4966 to construct an integrated genetic map that was 2,468.14 cM in length. These researchers used IM to test presence of putative QTLs and a multi-QTL model to declare QTLs and found 46 QTLs. There were 13 mapped QTLs for the tonnes of cane per hectare (TCH), 14 for sugar content in tonnes of sucrose per hectare (TSH), 11 for FIB and eight for POL%C. Singh et al. [34] studied the progeny of 207 individuals derived from a cross between cultivars Co86011 and CoH70, and they constructed two separate genetic maps, one for each parent. Through the CIM model, these researchers found 31 QTLs, with seven for BRIX and four for stalk number (SN). Thus, specific objectives must be taken into consideration for QTL mapping in sugarcane.

For QTL mapping in this study, we performed a permutation test to obtain the threshold for declaring significant QTLs [96]. The CIM model was a useful tool once it was able to identify regions with next QTL considering Araras and Ipaussu over the harvests (three years of evaluations). Seven QTLs were identified, being that a region located in LG4 at 43.32 cM showed QTLs for BRIX (B1-B3) and POL%C (P1-P2). The marker associated with the QTLs was identical for both traits, and the region that gave rise to this marker could be evaluated for future applications by sugarcane breeding programs. POL%C and BRIX are correlated traits [10], and although the commercial cultivars used as parents of the mapping population presented a small contrast in terms of phenotypic averages, especially for sucrose content, these results show that a combination of different alleles in each parent segregates and contributes to the observed variation in progeny. Furthermore, this result was expected because the parents of the mapping population are cultivars that were improved primarily to increase the sucrose content.

The percentage of phenotypic variation explained by each QTL ranged from 2.71% (FIB1) to 9.19% (B1) for Araras and from 5.38% (SD1) to 8.09% (P1) for Ipaussu. The sampling of the genome in single doses requires that QTL also segregate as single doses. Furthermore, the use of improved parents that have close phenotypic averages and that have some level of fixed alleles for traits of interest could decrease the chances of detecting QTLs with high rates of explained phenotypic variation. Nevertheless, the QTLs described here can be considered reliable because they have all taken into account the phenotypic average of three harvests and because one of them remained next between locations.

The common markers between locations could be regarded as potential regions to search genes that are involved in controlling quantitative traits. Because sugarcane is a clonally propagated crop when a strong marker-QTL association is detected in a full-sib progeny, it has an immediate impact on crop improvement via molecular marker selection because there is almost no further probability of crossover between the marker and the QTL [34]. Candidate genes for mapped QTLs could be inferred by homology analysis of sequences that originated from the associated markers. For the BRIX trait, we can highlight the homology of the marker SCSFAM1074E10_287, located in QTL B2, with *extended synaptotagmin-1-like*, which is a member of a membrane-trafficking protein family that is characterized by an N-terminal transmembrane region, a linker of variable size, and two C-terminal C2 domains in tandem [136]. C2 domains, identified as a conserved sequence motif in protein kinase C [137], are autonomously folded protein modules that generally act as calcium (Ca^{2+}) and phospholipid-binding domains and that were shown to represent autonomously folded Ca^{2+}-binding domains in synaptotagmins [138]. In addition, Ca^{2+} acts as a second messenger in the signal transduction pathways of hormones and environmental stimuli (touch, wind, chilling, light, and elicitors) [139], and several proteins that are involved in photosynthesis depend on Ca^{2+} [140]. Therefore, this homology could indicate some activity related to sugar transport between cells or organelles and mediated through Ca^{2+} sensors. Further studies are needed to expand this inference in pathways and regulatory networks for sugarcane.

For the FIB trait, a QTL (FIB1) showed different homology for each of the two adjacent markers and we can highlight the homology of the marker mf125302_409 with *zinc finger protein CONSTANS-LIKE 15*. The CONSTANS (CO) protein is a zinc finger transcription factor that contain two conserved domains (*i.e.*, a B-box zinc finger domain and a CCT [CO, CO-like, TOC1] domain), located in the region near the amino- and carboxy-terminus, respectively [145,146]. The CO proteins play a central role in the photoperiod pathway of *Arabidopsis* by mediating the circadian clock and floral integrators via positive regulation of FLOWERING LOCUS T expression [147-149]. In cotton (*Gossypium* spp.), which is the most important natural source of fiber for the textile industry, the CO5 protein (*CONSTANS-LIKE 5*) was up-regulated for cell wall modification and developing fibers in MD52ne, a near-isogenic line. Furthermore, using an F_2 population derived from a cross between MD52ne and MD90ne, stable QTLs for bundle fiber strength and fiber length were found, and CO5 was present on the QTL region for fiber length [150]. Therefore, although structural and organizational characterization of transposons detected by homology analysis is

necessary, the QTL region for the FIB trait could be associated with plant development and regulatory networks of cell wall formation in sugarcane.

5. Conclusions

Our understanding of the genetic architecture of traits of interest in sugarcane is increasing with the development of new analytical methods. The estimation of ploidy and allelic dosage through markers generated by GBS as well as the inclusion of these markers in an integrated genetic map of sugarcane were first observed in this study, and these markers showed great potential for QTL mapping. The CIM approach that provided additive and dominance effects and estimated the segregation patterns for all mapped QTLs was efficient for detecting possible stable QTLs among the evaluated locations. The verification of candidate genes for mapped QTLs showed great importance for new insights into the comprehensive relations between phenotypes and genotypes. It is still necessary to develop statistical approaches to enable the inclusion of markers at multiple doses to enhance the coverage by linking the SDMs that are pulverized by the genome. Moreover, QTL mapping with markers in multiple doses must be considered a major step in the understanding of regions that control quantitative traits in polyploid organisms and perhaps permit the verification of the allelic expression of phenotypic traits in the future.

6. Declarations

6.1 List of abbreviations

BRIX, soluble solid content; CCR, *cinnamoyl-CoA reductase*; CIM, composite interval mapping; COMT, *caffeic acid 3-O-methyltransferase*; DNA, deoxyribonucleic acid; FIB, fiber content; GBS, genotyping-by-sequencing; HGs, homo(eo)logous groups; IM, interval mapping; LGs, linkage groups; LOD, logarithm of the odds; MAP, maximum *a posteriori*; MAS, marker-assisted selection; MDM, multiple dose marker; NGS, next-generation sequencing; PCR, polymerase chain reaction; POL%C, sucrose content of cane; QTLs, quantitative trait loci; RNA, ribonucleic acid; SDMs, single-dose markers; SD, stalk diameter; SNPs, single nucleotide polymorphisms; SSRs, single sequence repeats, another term for microsatellites; SuPS, *sucrose phosphate synthase*; TRAP, target region amplification polymorphism; VCF, variant call format.

6.2 Ethics

Not applicable.

6.3 Consent to publish

Not applicable.

6.4 Competing interests

The authors declare that they have no competing interests.

6.5 Funding

This work was supported by grants from the INCT-Bioetanol (Instituto Nacional de Ciência e Tecnologia do Bioetanol), FINEP (Finaciadora de Estudos e Projetos) and FAPESP (Fundação de Amparo à Pesquisa de São Paulo, 08/52197-4). TWAB, GSP, MCM, EAC and CBCS received doctoral fellowships from FAPESP (10/50091-4, 12/25236-4, 10/50549-0, 10/50031-1 and 12/11109-0, respectively). COA received a doctoral fellowship from the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). FZB received a master's fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). APS and AAFG received research fellowships from CNPq.

6.6 Authors' contributions

MSC and HPH conceived and designed the field experiments. TWAB, MCM, GSP and COA performed the field experiment assessments and the phenotypic data analysis. TWAB and FZB performed the DNA extraction and SSR and TRAP marker data generation. MSC, AAFG, GRAM, FZB and TWAB designed the GBS experiments. GSP, AAFG and GRAM performed the GBS data analysis. APS, CBCS and EAC provided the RNA-seq data. TWAB, GSP, AAFG and GRAM performed the genetic map analysis. RG and AAFG performed the QTL mapping analysis. MSC, TWAB, GSP, AAFG, GRAM and APS participated in the study design and integrated the analysis of results. TWAB and GSP wrote

the manuscript draft, and MSC, AAFG, and APS edited and revised the manuscript. GRAM and RG critically read the manuscript. All the authors have read and approved the final manuscript.

6.7 Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

6.8 Acknowledgements

We gratefully acknowledge Marcelo Mollinari for sharing his experience working with SUPERMASSA software.

References

1. D'Hont A, Glaszmann JC. Sugarcane genome analysis with molecular markers, a first decade of research. Proceedings of the International Society for Sugar Cane Technologies. 2001;24:556-9.
2. D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC. Determination of basic chromosome numbers in the genus Saccharum by physical mapping of ribosomal RNA genes. Genome. 1998;41:221-5.
3. D'Hont A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet Genome Res. 2005;109:27-33.
4. Palhares AC, Rodrigues-Morais TB, Van Sluys MA, Domingues DS, Maccheroni W, Jordão H, et al. A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. BMC Genet. 2012;13:51.
5. Piperidis G, Piperidis N, D'Hont A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Genet Genomics. 2010;284:65-73.
6. Ha S, Moore PH, Heinz D, Kato S, Ohmido N, Fukui K. Quantitative chromosome map of the polyploid Saccharum spontaneum by multicolor fluorescence in situ hybridization and imaging methods. Plant Mol Biol. 1999;39:1165-73.

7. Daniels J, Roach BT. Taxonomy and evolution. In: Heinz DJ, editor. Sugarcane improvement through breeding. Amsterdam: Elsevier Press; 1987. p. 7–84.
8. Irvine JE. Saccharum species as horticultural classes. Theor Appl Genet. 1999;98:186-94.
9. Roach BT. Cytological studies in Saccharum. Chromosome transmission in interspecific and intergeneric crosses. Proceedings of the International Society for Sugar Cane Technologies. 1969;13:901-20.
10. Balsalobre T, Mancini MC, Pereira GS, Anoni CO, Barreto FZ, Hoffmann HP, et al. A mixed-model approach for analysis of yield components and brown rust resistance in full-sib families of sugarcane. Agron J. 2016;108:1-14. doi:10.2134/agronj2015.0430.
11. Eksteen A, Singels A, Ngxaliwe S. Water relations of two contrasting sugarcane genotypes. Field Crops Res. 2014;168:86-100.
12. Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnol J. 2010;8:263-76. doi: 10.1111/j.1467-7652.2009.00491.x.
13. Adams KL, Wendel JF. Polyploidy and genome evolution in plants. Curr Opin Plant Biol. 2005;8:135-41.
14. Dubcovsky J, Dvorak J. Genome plasticity a key factor in the success of polyploid wheat under domestication. Science. 2007;316:1862-6.
15. te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesová M, et al. The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. 2012;109:19–45.
16. Madlung A. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity. 2013;110:99-104.
17. Yagi M, Yamamoto T, Isobe S, Hirakawa H, Tabata S, Tanase K, et al. Construction of a reference genetic linkage map for carnation (dianthus caryophyllus L.). BMC Genomics. 2013;14:734.
18. Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, et al. High-resolution genetic maps of eucalyptus improve eucalyptus grandis genome assembly. New Phytol. 2015;206:1283–96.
19. Deokar AA, Ramsay L, Sharpe AG, Diapari M, Sindhu A, Bett K, et al. Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. BMC Genomics. 2014;15:708.

20. Portis E, Mauromicale G, Mauro R, Acquadro A, Scaglione D, Lanteri S. Construction of a reference molecular linkage map of globe artichoke (Cynara cardunculus var. scolymus). Theor Appl Genet. 2009;120:59-70.
21. Hudson CJ, Freeman JS, Kullan AR, Petroli CD, Sansaloni CP, Kilian A, et al. A reference linkage map for eucalyptus. BMC Genomics. 2012;13:240.
22. Hong Y, Chen X, Liang X, Liu H, Zhou G, Li S, et al. A SSR-based composite genetic linkage map for the cultivated peanut (Arachis hypogaea L.) genome. BMC Plant Biol. 2010;10:17.
23. Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genet. 1992;83:294–300.
24. Grattapaglia D, Sederoff R. Genetic linkage maps of eucalyptus grandis and eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genet. 1994;137:1121-37.
25. Grivet L, D'Hont A, Dufour P, Hamon P, Roques D, Glaszmann JC. Comparative genome mapping of sugar cane with other species within the Andropogoneae tribe. Hered. 1994;73:500-8.
26. Da Silva J, Honeycutt RJ, Burnquist W, Al-Janabi SM, Sorrells ME, Tanksley SD, et al. Saccharum spontaneum L. 'SES 208' genetic linkage map combining RFLP and PCR based markers. Mol Breeding. 1995;1:165-79.
27. Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, et al. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet. 1997;94:409-18.
28. Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, et al. Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics. 1998;150:1663-82.
29. Asnaghi C, Paulet F, Kaye C, Grivet L, Deu M, Glaszmann JC, et al. Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. Theor Appl Genet. 2000;101:962-9.
30. Edmé SJ, Glynn NG, Comstock JC. Genetic segregation of microsatellite markers in Saccharum officinarum and S Spontaneum. Heredity. 2006;97:366-75.
31. Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GR, Oliveira KM, et al. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor Appl Genet. 2012;124:835-49.

32. Pastina MM, Pinto LR, Oliveira KM, Souza AP, Garcia AAF. Molecular mapping of complex traits. In: Henry RJ, Kole C, editors. Genetics, genomics and breeding of sugarcane. Boca Raton, FL: CRC Press; 2010.

33. Andru S, Pan Y-B, Thongthawee S, Burner DM, Kimbeng CA. Genetic analysis of the sugarcane (Saccharum spp.) cultivar 'LCP 85–384'. I. Linkage mapping using AFLP, SSR, and TRAP markers. Theor Appl Genet. 2011;123:77-93.

34. Singh RK, Singh SP, Tiwari DK, Srivastava S, Singh SB, Sharma ML, et al. Genetic mapping and QTL analysis for sugar yield-related traits in sugarcane. Euphytica. 2013;191:333-53.

35. Aitken KS, McNeil MD, Hermann S, Bundock PC, Kilian A, Heller-Uszynska K, et al. A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput diversity array technology (DArT) markers. BMC Genomics. 2014;15:152.

36. Wu P, Liao CY, Hu B, Yi KK, Jin WZ, Ni JJ, et al. QTLs and epistasis for aluminum tolerance in rice (Oryza sativa L.) at different seedling stages. Theor Appl Genet. 2000;100:1295-303.

37. Garcia AA, Kido EA, Meza AN, Souza HM, Pinto LR, Pastina MM, et al. Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet. 2006;112:298-314.

38. Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, et al. Functional integrated genetic linkage map based on EST-markers for a sugarcane (Saccharum spp.) commercial cross. Mol Breed. 2007;20:189-208.

39. Garcia AA, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira ML, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep. 2013;3:3399. doi: 10.1038/srep03399.

40. Wang J, Roe B, Macmil S, Yu Q, Murray JE, Tang H, et al. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. BMC Genomics. 2010;11:261.

41. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499-510.

42. He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci. 2014;5:484. doi: 10.3389/fpls.2014.00484.

43. Kim SR, Ramos J, Ashikari M, Virk PS, Torres EA, Nissila E, et al. Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, Oryza sativa L. Rice. 2016;9:12. doi: 10.1186/s12284-016-0084-7.

44. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 2007;17:240-8. doi: 10.1101/gr.5681207.

45. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS ONE. 2011;6:e19379.

46. Huang Y-F, Poland JA, Wight CP, Jackson EW, Tinker NA. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLOS ONE. 2014;9:e102448. doi: 10.1371/journal.pone.0102448.

47. Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Valè G, et al. Next generation breeding. Plant Sci. 2016;242:3-13. doi: 10.1016/j.plantsci.2015.07.010.

48. Byrne S, Czaban A, Studer B, Panitz F, Bendixen C, Asp T. Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. PLOS ONE. 2013;8:e57438. doi: 10.1371/journal.pone.0057438.

49. Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLOS ONE. 2013;8:e54603. doi: 10.1371/journal.pone.0054603.

50. Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, et al. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 (Bethesda). 2013;3:1903–26. doi: 10.1534/g3.113.008227.

51. Heslot N, Rutkoski J, Poland J, Jannink J-L, Sorrells ME. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLOS ONE. 2013;8:e74612. doi: 10.1371/journal.pone.0074612.

52. Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, et al. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP

markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet. 2013;126:2699-716. doi: 10.1007/s00122-013-2166-x.

53. Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, et al. An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genomics. 2014;15:104. doi: 10.1186/1471-2164-15-104.
54. Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S. High-density linkage map construction and mapping of seed trait QTLs in chickpea (Cicer arietinum L.) using genotyping-by-sequencing (GBS). Sci Rep. 2015;5:17512. doi: 10.1038/srep17512.
55. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 2002;5:94-100.
56. Hackett CA, McLean K, Bryan GJ. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLOS ONE. 2013;8:e63939. doi: 10.1371/journal.pone.0063939.
57. Lee J, Izzah NK, Jayakodi M, Perumal S, Joh HJ, Lee HJ, et al. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. BMC Plant Biol. 2015;15:32. doi: 10.1186/s12870-015-0424-6.
58. Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR. A comparison of analysis methods for late-stage variety evaluation trials. Aust NZ. J Stat. 2010;52:125–49. doi: 10.1111/j.1467-842X.2010.00570.x.
59. Aitken KS, Hermann S, Karno K, Bonnett GD, McIntyre LC, Jackson PA. Genetic control of yield related stalk traits in sugarcane. Theor Appl Genet. 2008;117:1191-203.
60. Margarido GRA, Pastina MM, Souza AP, Garcia AAF. Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. Mol Breeding. 2015;35:175.
61. Gazaffi R, Margarido GRA, Pastina MM, Mollinari M, Garcia AAF. A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. Tree Genet Genomes. 2014;10:791-801.
62. Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, et al. QTL mapping of growth-related traits in a full-sib family of rubber tree (hevea brasiliensis) evaluated in a sub-tropical climate. PLOS ONE. 2013;8:e61238.
63. Zeng ZB. Precision mapping of quantitative trait loci. Genet. 1994;136:1457–68.

64. Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res. 2003;13:2725-35.

65. Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TW, Canesin LE, Pinto LR, et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLOS ONE. 2014;9:e88462.

66. Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, et al. The sugarcane genome challenge: strategies for sequencing a highly complex genome. Trop Plant Biol. 2011;4:145–56.

67. de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GM, Del Bem LE, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics. 2014;15:540. doi: 10.1186/1471-2164-15-540.

68. Metcalfe CJ, Oliveira SG, Gaiarsa JW, Aitken KS, Carneiro MS, Zatti F, et al. Using quantitative PCR with retrotransposon-based insertion polymorphisms as markers in sugarcane. J Exp Bot. 2015;66:4239-50. doi: 10.1093/jxb/erv283.

69. Al-Janabi SM, Forget L, Dookun A. An improved and rapid protocol for the isolation of polysaccharide-and polyphenol-free sugarcane DNA. Plant Mol Biol Rep. 1999;17:1-8.

70. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. Tassel-GBS: a high capacity genotyping by sequencing analysis pipeline. PLOS ONE. 2014;9:e90346.

71. Grativol C, Regulski M, Bertalan M, McCombie WR, da Silva FR, Zerlotini Neto A, et al. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. Plant J. 2014;79:162-72.

72. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The sorghum bicolor genome and the diversification of grasses. Nature. 2009;457:551-6.

73. Serang O, Mollinari M, Garcia AA. Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLOS ONE. 2012;7:e30906. doi: 10.1371/journal.pone.0030906.

74. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811–2.

75. Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, et al. Characterization of new polymorphic functional markers for sugarcane. Genome. 2009;52:191-209.

76. Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, de Souza AP. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed. 2006;125:378-84.

77. Marconi TG, Costa EA, Miranda HR, Mancini MC, Cardoso-Silva CB, Oliveira KM, et al. Functional markers for gene mapping and genetic diversity studies in sugarcane. BMC Res Notes. 2011;4:264.

78. Singh RK, Jena SN, Khan S, Yadav S, Banarjee N, Raghuvanshi S, et al. Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. Gene. 2013;524:309-29.

79. Cordeiro GM, Taylor GO, Henry RJ. Characterisation of microsatellite markers from sugarcane (Saccharum spp.), a highly polyploid species. Plant Sci. 2000;155:161-8.

80. Raboin L-M, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, et al. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. Theor Appl Genet. 2006;112:1382-91.

81. Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, et al. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [sorghum bicolor (L.) Moench]. Theor Appl Genet. 1996;93:190-8.

82. Kong L, Dong J, Hart GE. Characteristics, linkage-map positions, and allelic differentiation of sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet. 2000;101:438-48.

83. Wang ML, Wang ML, Barkley NA, Yu J-, Dean RE, Newman ML, et al. Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. Plant Genet Resour Newsl. 2005;3:45–57.

84. Li G, Quiros CF. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction : its application to mapping and gene tagging in Brassica. Theor Appl Genet. 2001;103:455-61.

85. Alwala S, Suman A, Arro JA, Veremis JC, Kimbeng CA. Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 2006;46:448-55.

86. Creste S, Sansoli DM, Tardiani ACS, Silva DN, Gonçalves FK, Fávero TM, et al. Comparison of AFLP, TRAP and SSRs in the estimation of genetic relationships in sugarcane. Sugar Tech. 2010;12:150-54.

87. Suman A, Ali K, Arro J, Parco AS, Kimbeng CA, Baisakh N. Molecular diversity among members of the Saccharum complex assessed using TRAP Markers based on lignin-related genes. Bioenerg Res. 2012;5:197–205.

88. Hu J, Vick BA. Target region amplification polymorphism: a novel marker technique for plant genotyping. Plant Mol Biol Rep. 2003;21:289–94.

89. Creste S, Neto AT, Figueira A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep. 2001;19:299–306.

90. Margarido GR, Souza AP, Garcia AA. OneMap: software for genetic mapping in outcrossing species. Hereditas. 2007;144:78-9.

91. Wu R, Ma C-X, Painter I, Zeng Z-B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor Popul Biol. 2002;61:349–63.

92. Wu R, Ma C-X, Wu SS, Zeng Z-B. Linkage mapping of sex-specific differences. Genet Res. 2002;79:85–96.

93. Jiang C, Zeng Z-B. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica. 1997;101:47-58.

94. Voorrips RE. Computer Note MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 1994;93:77–8.

95. Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;121:185-199.

96. Chen L, Storey JD. Relaxed significance criteria for linkage analysis. Genetics. 2006;173:2371–81.

97. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. http://www.R-project.org/.

98. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6. doi: 10.1093/bioinformatics/bti610.

99. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178–86.

100. Li H, Vikram P, Singh RP, Kilian A, Carling J, Song J, et al. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genomics. 2015;16:216.

101. Heffelfinger C, Fragoso CA, Moreno MA, Overton JD, Mottinger JP, Zhao H, et al. Flexible and scalable genotyping-by-sequencing strategies for population studies. BMC Genomics. 2014;15:979. doi: 10.1186/1471-2164-15-979.

102. Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, et al. Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics. 2013;193:1073-81. doi: 10.1534/genetics.112.147710.

103. Jiang Z, Wang H, Michal JJ, Zhou X, Liu B, Woods LC, et al. Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. Int J Biol Sci. 2016;12:100-8. doi: 10.7150/ijbs.13498.

104. Grivet L, Arruda P. Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol. 2002;5:122-7.

105. Mollinari M, Serang O. Quantitative SNP genotyping of polyploids with MassARRAY and other platforms. Methods Mol Biol. 2015;1245:215–41. doi: 10.1007/978-1-4939-1966-6_17.

106. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012;13:36-46.

107. Reffay N, Jackson PA, Aitken KS, Hoarau JY, D'hont A, Besse P, et al. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol Breed. 2005;15:367-81.

108. Aitken KS, Jackson PA, McIntyre CL. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet. 2005;110:789-801.

109. Aitken KS, Jackson PA, McIntyre CL. Construction of a genetic linkage map for Saccharum officinarum incorporating both simplex and duplex markers to increase genome coverage. Genome. 2007;50:742–56.

110. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in

hexaploid wheat landraces and cultivars. Proc Natl Acad Sci U S A. 2013;110:8057-62.

111. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLOS ONE. 2012;7:e32253. doi: 10.1371/journal.pone.0032253.
112. Wang N, Li F, Chen B, Xu K, Yan G, Qiao J, et al. Genome-wide investigation of genetic changes during modern breeding of Brassica napus. Theor Appl Genet. 2014;127:1817-29.
113. Butterfield MK, D'Hont A, Berding N. The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc. South African Sugar Technology Assoc. 2001;75:1-5.
114. Aitken KS, McNeil MD, Berkman PJ, Hermann S, Kilian A, Bundock PC, et al. Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. BMC Plant Biol. 2014;14:190.
115. Glaszmann JC, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, et al. Comparative genome analysis between several tropical grasses. Euphytica. 1997;96:13-21.
116. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008;320:486-8.
117. Salse J, Abrouk M, Murat F, Quraishi UM, Feuillet C. Improved criteria and comparative genomics tool provide new insights into grass paleogenomics. Brief Bioinform. 2009;10:619-30.
118. Margarido GRA, Heckermann D. ConPADE: Genome Assembly Ploidy Estimation from next-generation sequencing data. PLOS Comput Biol. 2015: 11(4): e1004229. doi:10.1371/journal.pcbi.1004229.
119. Malosetti M, Ribaut J-M, van Eeuwijk FA. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol. 2013;4:44.
120. Al-Janabi SM, Parmessur Y, Kross H, Dhayan S, Saumtally S, Ramdoyal K, et al. Identification of a major quantitative trait locus (QTL) for yellow spot (Mycovellosiella koepkei) disease resistance in sugarcane. Mol Breed. 2007;19:1-14.
121. Sills GR, Bridges W, Al-Janabi SM, Sobral BWS. Genetic analysis of agronomic traits in a cross between sugarcane (Saccharum officinarum L.) and its presumed progenitor. (S robustum Brandes & Jesw. Ex Grassl). Mol. Breed. 1995;1:355-63.

122. Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, et al. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570. Theor Appl Genet. 1996;92:1059-64.

123. Guimarães CT, Sills GR, Sobral BW. Comparativemapping of Andropogoneae: Saccharum L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci U S A. 1997;94:14261-6.

124. Asnaghi C, D'hont A, Glaszmann JC, Rott P. Resistance of sugarcane cultivar R570 to Puccinia melanocephala from different geographic locations. Plant Dis. 2001;85:282–6.

125. Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 2001;11:2075-84.

126. Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH. Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome. 2002;45:794-803.

127. Ming R, Wang W, Draye X, Moore H, Irvine E, Paterson H. Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. Theor Appl Genet. 2002;105:332-45.

128. Hoarau JY, Grivet L, Offmann B, Raboin LM, Diorflar JP, Payet J, et al. Genetic dissection of a modern sugarcane cultivar (Saccharum spp.).II. Detection of QTLs for yield components. Theor Appl Genet. 2002;105:1027-37.

129. Silva JAd, Bressiani JA. Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny. Genet Mol Biol. 2005;28:294-8.

130. Aitken KS, Jackson PA, McIntyre CL. Quantitative trait loci identified for sugar related traits in a sugarcane (Saccharum spp.) cultivar x Saccharum officinarum population. Theor Appl Genet. 2006;112:1306-17.

131. Raboin LM, Pauquet J, Butterfield M, D'hont A, Glaszmann JC. Analysis of genome-wide linkage disequilibriumin the highly polyploid sugarcane. Theor Appl Genet. 2008;116:701-14.

132. Wei X, Jackson PA, McIntyre CL, Aitken KS, Croft B. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. Theor Appl Genet. 2006;114:155-64.

133. Piperidis N, Jackson PA, D'hont A, Besse P, Hoarau J, Courtois B, et al. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. Mol Breed. 2008;21:233-47.

134. Pinto LR, Garcia AAF, Pastina MM, Teixeira LHM, Bressiani JA, Ulian EC, et al. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (Saccharum spp.) commercial cross. Euphytica. 2010;172:313-27.

135. Jordan DR, Casu RE, Besse P, Carroll BC, Berding N, McIntyre CL. Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. Genome. 2004;47:988-93.

136. Craxton M. Synaptotagmin gene content of the sequencedgenomes. BMC Genomics. 2004;5:43.

137. Coussens L, Parker PJ, Rhee L, Yang-Feng TL, Chen E, Waterfield MD, et al. Multiple, distinct forms of bovine and human protein kinase C suggest diversity in cellular signaling pathways. Science. 1986;233:859–66.

138. Schapire AL, Voigt B, Jasik J, Rosado A, Lopez-Cobollo R, Menzel D, et al. Arabidopsis Synaptotagmin 1 is required for the maintenance of plasma membrane integrity and cell viability. Plant Cell. 2008;20:3374–88.

139. Shin DH, Choi M-G, Lee HK, Cho M, Choi S-B, Choi G, et al. Calcium dependent sucrose uptake links sugar signaling to anthocyanin biosynthesis in Arabidopsis. Biochem Biophys Res Commun. 2013;430:634–9.

140. Hochmal AK, Schulze S, Trompelt K, Hippler M. Calcium-dependent regulation of photosynthesis. Biochim Biophys Acta. 2015;1847:993–1003.

141. Manetti ME, Rossi M, Cruz GMQ, Saccaro NL, Nakabashi M, Altebarmakian V, et al. Mutator system derivatives isolated from sugarcane genome sequence. Trop Plant Biol. 2012;5:233-43.

142. Saccaro NL, Van Sluys M-A, de Mello Varani A, Rossi M. Mudra-like sequences from rice and sugarcane cluster as two bona fide transposon clades and two domesticated transposases. Gene. 2007;392:117-25.

143. de Jesus EM, Cruz EA, Cruz GM, Van Sluys MA. Diversification of hAT transposase paralogues in the sugarcane genome. Mol Genet Genomics. 2012;287:205-19.

144. Bundock P, Hooykaas P. An Arabidopsis hat-like transposase is essential for plant development. Nature. 2005;436:282-4.

145. Robson F, Costa MM, Hepworth SR, Vizir I, Piñeiro M, Reeves PH, et al. Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J. 2001;28:619–31.

146. Chou ML, Shih MC, Chan MT, Liao SY, Hsu CT, Haung YT, et al. Global transcriptome analysis and identification of a CONSTANS-like gene family in the orchid Erycina pusilla. Planta. 2013;237:1425–41.

147. Turck F, Fornara F, Coupland G. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol. 2008;59:573–94.

148. Kovi MR, Sablok G, Bai X, Wendell M, Rognli OA, Yu H, et al. Expression patterns of photoperiod and temperature regulated heading date genes in Oryza sativa. Comput Biol Chem. 2013;45:36–41.

149. Liu T, Zhu S, Tang Q, Tang S. Identification of a CONSTANS homologous gene with distinct diurnal expression patterns in varied photoperiods in ramie (Boehmeria nivea L Gaud). Gene. 2015;560:63–70.

150. Islam MS, Fang DD, Thyssen GN, Delhom CD, Liu Y, Kim HJ. Comparative fiber property and transcriptome analyses reveal key genes potentially related to high fiber strength in cotton (Gossypium hirsutum L.) line MD52ne. BMC Plant Biol. 2016;16:36. doi: 10.1186/s12870-016-0727-2.

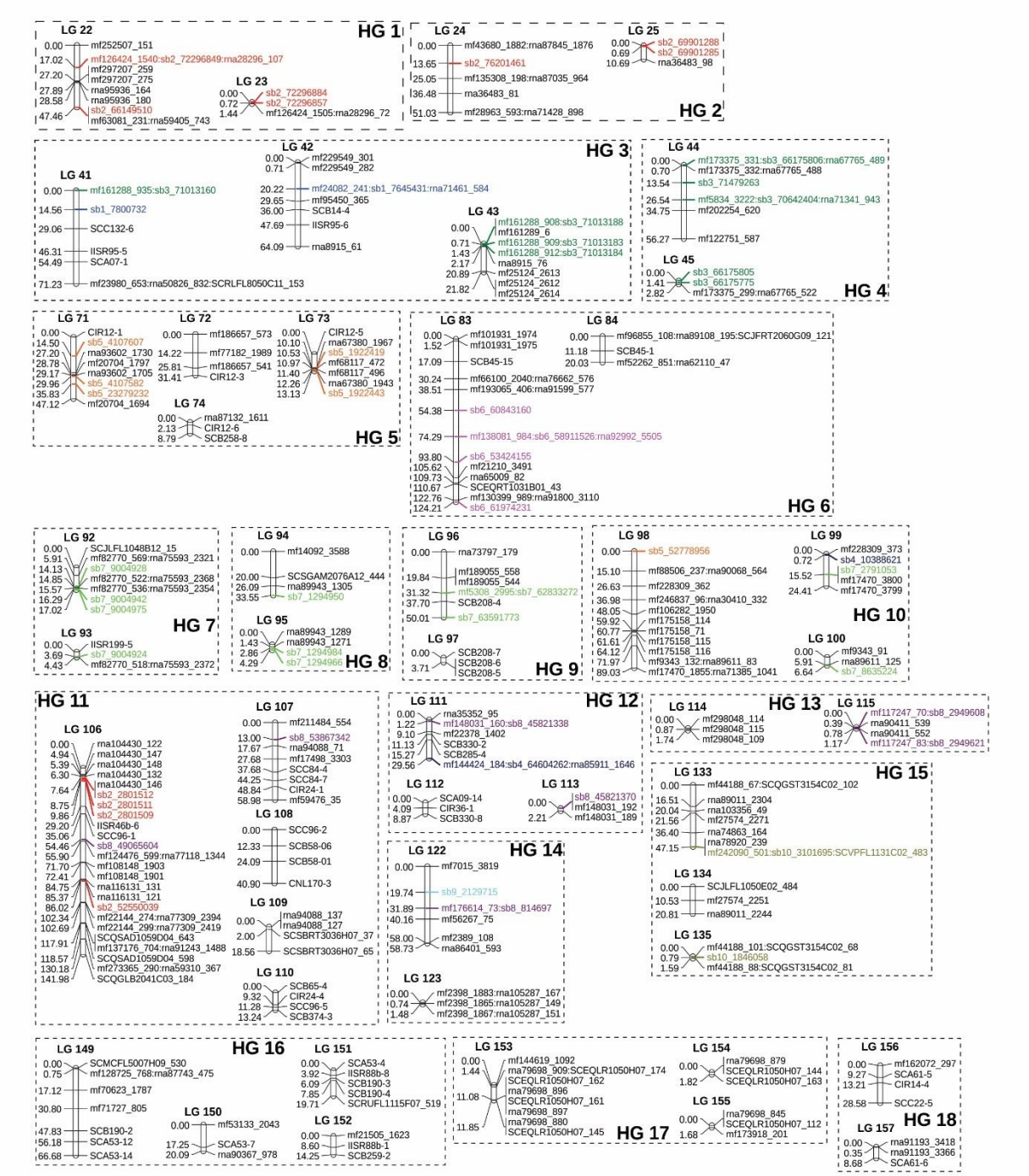
Supplementary Material

Supplementary Table 1 - BOWTIE2 alignment results of 3,103,708 million GBS tags in absolute and relative (in parentheses) values.

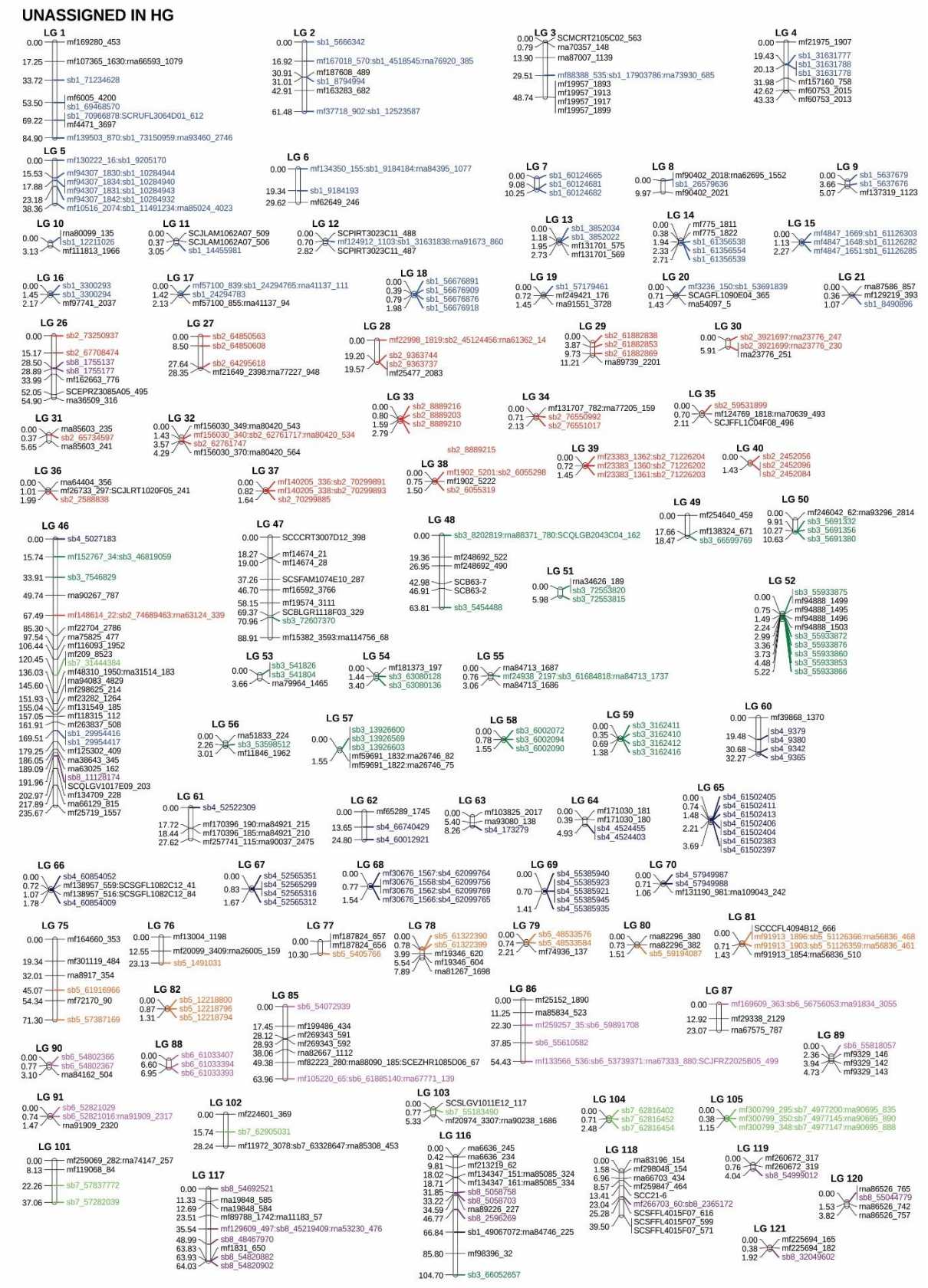
GBS-TASSEL pipeline pseudo-references	Non-aligned tags	Aligned tags		
		Overall alignment	Unique alignment	Non-unique alignment
Methyl-filtered sugarcane genome	374,251 (12.06%)	2,729,457 (87.94%)	1,823,623 (58.76%)	905,834 (29.18%)
Sorghum bicolor genome (v. 2.1)	1,791,047 (57.71%)	1,312,661 (42.29%)	1,060,705 (34.17%)	251,956 (8.12%)
RNA-seq sugarcane transcriptome	1,907,985 (61.47%)	1,195,723 (38.53%)	1,099,697 (35.43%)	96,026 (3.10%)
SUCEST project sequences	2,362,171 (76.11%)	741,537 (23.89%)	256,450 (8.26%)	485,087 (15.63%)

Supplementary Table 2 – Markers selected as cofactors for mapping through the composite interval mapping (CIM) model for the soluble solid content (BRIX, in °Brix), sucrose content of cane (POL%C, in %), fiber content (FIB, in %) and stalk diameter (SD, in mm) traits for the locations Araras-SP and Ipaussu-SP.

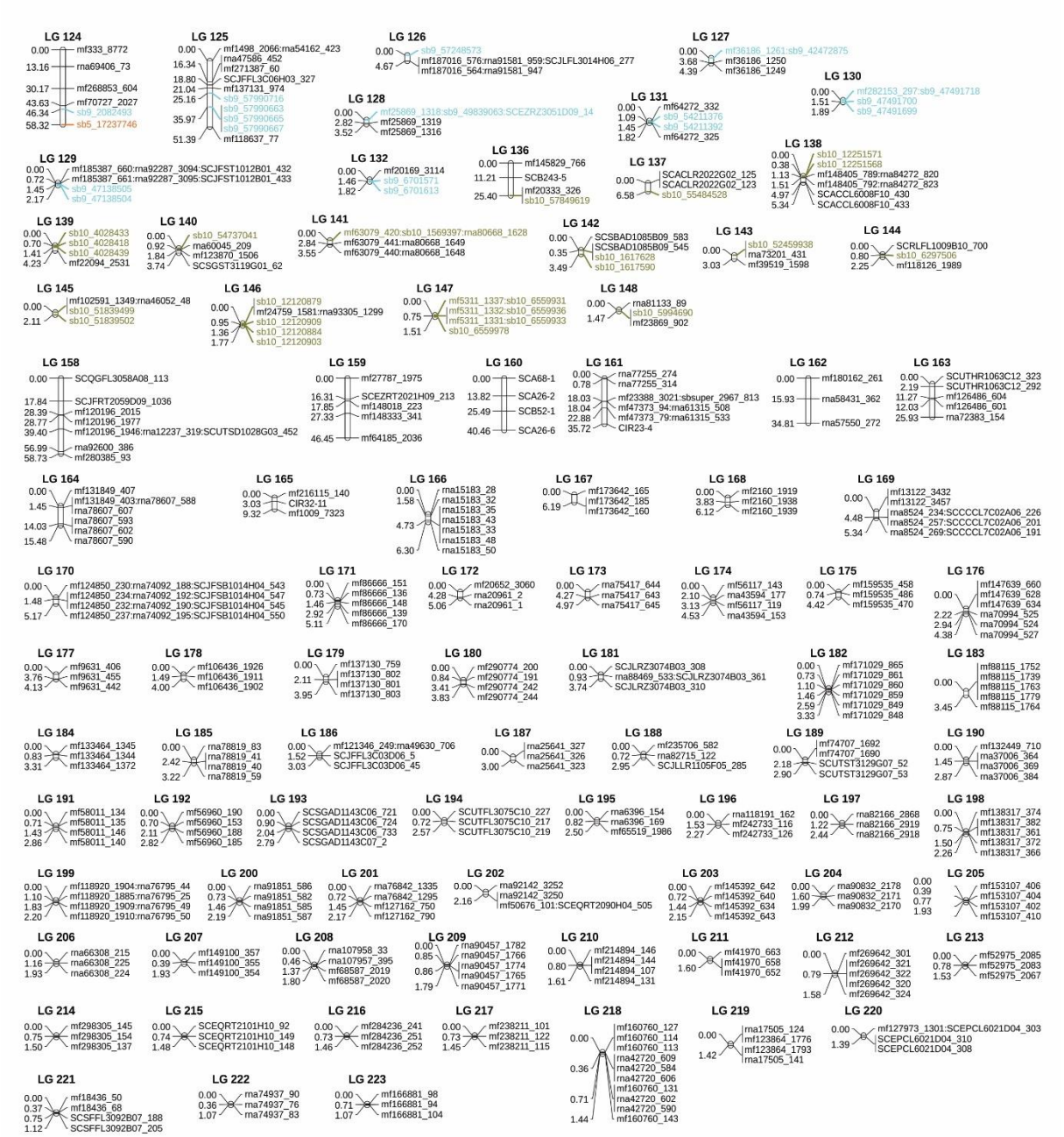
Location	Trait	LG	Marker selected as cofactors for the CIM model	Location	Trait	LG	Marker selected as cofactors for the CIM model
Araras	BRIX	4	mf60753_2013	Araras	FIB	122	sb9_2129715
Araras	BRIX	6	mf62649_246	Araras	FIB	124	mf333_8772
Araras	BRIX	26	sb8_1755137	Araras	FIB	167	mf173642_165
Araras	BRIX	42	mf229549_282	Ipaussu	FIB	47	sb3_72607370
Araras	BRIX	47	SCSFAM1074E10_287	Ipaussu	FIB	54	sb3_63080136
Ipaussu	BRIX	4	mf60753_2013	Ipaussu	FIB	60	mf39868_1370
Ipaussu	BRIX	6	mf62649_246	Ipaussu	FIB	124	mf333_8772
Ipaussu	BRIX	26	sb8_1755137	Ipaussu	FIB	167	mf173642_165
Ipaussu	BRIX	42	mf229549_282	Araras	SD	7	sb1_60124665
Ipaussu	BRIX	46	mf209_8523	Araras	SD	35	SCJFFL1C04F08_496
Ipaussu	BRIX	98	mf17470_1855:rna71385_1041	Araras	SD	106	rna104430_132
Araras	POL%C	4	mf60753_2013	Araras	SD	133	rna74863_164
Araras	POL%C	6	mf62649_246	Araras	SD	146	sb10_12120909
Araras	POL%C	26	sb8_1755137	Araras	SD	150	rna90367_978
Araras	POL%C	42	mf229549_282	Araras	SD	174	mf56117_119
Araras	POL%C	51	sb3_72553815	Araras	SD	218	rna42720_590
Ipaussu	POL%C	4	mf60753_2013	Ipaussu	SD	7	sb1_60124665
Ipaussu	POL%C	6	mf62649_246	Ipaussu	SD	13	sb1_3852034
Ipaussu	POL%C	26	sb8_1755137	Ipaussu	SD	35	SCJFFL1C04F08_496
Ipaussu	POL%C	42	mf229549_282	Ipaussu	SD	96	rna73797_179
Ipaussu	POL%C	51	sb3_72553815	Ipaussu	SD	106	rna104430_132
Araras	FIB	46	rna90267_787	Ipaussu	SD	133	rna74863_164
Araras	FIB	47	sb3_72607370	Ipaussu	SD	150	rna90367_978
Araras	FIB	60	mf39868_1370	Ipaussu	SD	174	rna43594_177



Supplementary Figure 1 - Genetic map of sugarcane that was generated with a population of 151 full sibs that originated from a commercial cross between the cultivars SP80-3280 and RB835486. The linkage groups (LGs) were separated into 18 homo(eo)logous groups (HGs). The numbers on the left of the LGs are the cumulative genetic distances in Kosambi centimorgans. The marker names are shown on the right. Markers corresponding to each sorghum chromosome are highlighted in different colors.



Supplementary Figure 1 - Continued.



Supplementary Figure 1 - Continued.

Este tópico foi adicionado para melhor entendimento da análise de redundância entre marcadores baseados em GBS e da análise de construção do mapa genético de ligação, ambas apresentadas no Capítulo 2.

1. *Análise de redundância entre marcadores baseados em GBS*

Após os marcadores baseados em GBS passarem pela análise no software SUPERMASSA para estimativa da ploidia e da dosagem alélica, àqueles em dose única e com ploidias entre 6 e 14 foram inspecionados para redundância dentro de cada referência e entre referências baseado na chamada dos genótipos, ou seja, locos com padrões de segregação duplicados dentro e entre referências tiveram uma única chamada levada para análise de ligação, sendo armazenada a relativa informação de redundância. Um exemplo da verificação de redundância pode ser vista na Tabela 1.

Tabela 1 - Chamada de 10 genótipos da população de mapeamento para seis marcadores baseados em GBS (*Genotyping-by Sequecing*) identificados com base nas quatro pseudo-referências: genoma do *Sorghum bicolor* (sb), genoma metil-filtrado da cana-de-açúcar (mf), transcriptoma da cana-de-açúcar RNA-seq (rna) e sequências dos projeto SUCEST (SC).

	Genótipos									
Marcadores	1	2	3	4	5	6	7	8	9	10
sb1_1	*a*	*b*	*b*	*ab*	*a*	*a*	*ab*	*ab*	*a*	*b*
sb8_3	*a*	*b*	*b*	*ab*	*a*	*a*	*ab*	*ab*	*a*	*b*
mf12_1	*ab*	*b*	*a*	*ab*	*b*	*a*	*a*	*ab*	*a*	*a*
rna31_4	*ab*	*b*	*a*	*ab*	*b*	*a*	*a*	*ab*	*a*	*a*
SC3_6	*ab*	*b*	*a*	*ab*	*b*	*a*	*a*	*ab*	*a*	*a*
mf25_9	*a*	*ab*	*b*	*ab*	*a*	*b*	*ab*	*a*	*a*	*ab*

No exemplo, apenas o marcador mf25_9 não apresentou redundância nos genótipos avaliados. Os marcadores sb1_1 e sb8_3 tiveram o mesmo padrão de segregação e devem ter apenas uma única chamada levada para a análise de ligação; sb1_1:sb8_3. Da mesma forma, foram redundantes os marcadores mf12_1, rna31_4 e SC3_6, e também devem ter uma única chamada na análise de ligação: mf12_1:rna31_4:SC3_6.

Apesar dos marcadores apresentarem posições físicas diferentes (por exemplo, sb1_1 e sb8_3 ancorados no cromossomo 1 e 8, respectivamente, de *Sorghum*

bicolor), na análise de ligação estes marcadores estariam sobrepostos na mesma posição do mapa genético.

2. Construção do mapa genético de ligação

Através dos marcadores moleculares em doses únicas oriundos de GBS, de SSR e de TRAP, avaliados na população de mapeamento originada a partir do cruzamento entre os cultivares SP80-3280 e RB835486, foram gerados cinco mapas genéticos de ligação integrados (Tabela 2) utilizando o software Onemap (v. 2.0-4) e os comandos citados em Material e métodos, item "*2.4 Linkage map construction and homo(eo)logous group assignment*", do Capítulo 2. O Mapa E foi utilizado para o mapeamento de QTLs por ter a melhor densidade de marcadores e pelos *heatmaps* dos grupos de ligação mostrarem bom ordenamento dos marcadores.

Tabela 2 - Número de grupos de ligação, número de marcadores mapeados, tamanho total do mapa em centimorgans (cM) e densidade de marcadores dos cinco mapas genéticos de ligação integrados gerados a partir dos marcadores moleculares oriundos de GBS (*Genotyping-by-Sequecing*), microssatélites (*Simple Sequence Repeats*, SSR) e TRAP (*Target Region Amplification Polymorphism*) avaliados na progênie do cruzamento entre os cultivares SP80-3280 e RB835486.

	Mapa A	Mapa B	Mapa C	Mapa D	Mapa E
Número de grupos de ligação	500	580	609	330	223
Número de marcadores mapeados	2,293	1,968	1,834	1,276	993
Tamanho total do mapa (cM)	16,485.51	10,178.01	7,657.37	5,928.09	3,682.04
Densidade de marcadores	7.19	5.17	4.17	4.64	3.70

Os Mapas A, B, C e D foram construídos utilizando LOD score > 9.0 e fração de recombinação < 0.15 e o Mapa E com LOD score > 9.0 e fração de recombinação < 0.10. Em adição, os Mapas A, B e C tiveram excluídos grupos de ligação menores que 1 centimorgan (cM), enquanto que os Mapas D e E tiveram excluídos grupos de ligação menores que 1 cM e com duas marcadores.

Um grupo de ligação com bom ordenamento de marcadores apresenta *heatmap* com cores quentes ao longo da diagonal e cores frias nos extremos (Margarido et al., 2007). Um exemplo de bom ordenamento pode ser visto na Figura 1. Assim, com a inspeção visual dos *heatmaps* de cada grupo de ligação foi possível melhorar o ordenamento dos marcadores e, consequentemente, a densidade de marcadores.

A técnica de GBS juntamente com as estimativas de ploidia e dosagem alélica, realizadas através do software SUPERMASSA, permitiram a utilização de marcadores com segregação 1:2:1 na construção do mapa genético. Estes marcadores são importantes para construção de mapas genéticos integrados, assim como os marcadores com segregação 3:1 obtidos a partir de SSR e TRAP. No entanto, aumentando a quantidade dessas classes de segregação, também aumentamos as chances de obtenção de grandes grupos de ligação, os quais apresentam dificuldade para ordenamento dos marcadores e precisam ser quebrados em grupos menores (Figura 2).

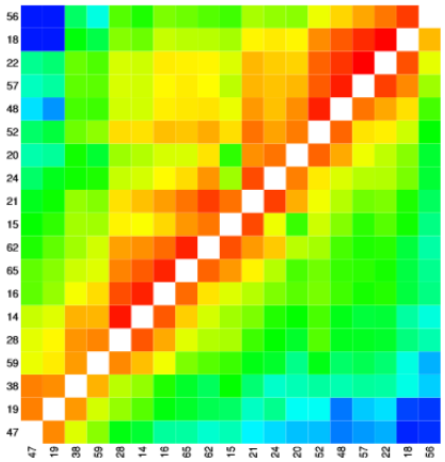


Figura 1 - *Heatmap* com cores quentes ao longo da diagonal e cores frias nos extremos indicando bom ordenamento dos marcadores dentro de um grupo de ligação (Margarido *et al*., 2007). As cores quentes acima e abaixo da diagonal representam maiores valores de LOD score e menores valores de fração de recombinação, respectivamente.

Desta forma, um exemplo do procedimento realizado em cada grupo de ligação do Mapa A até chegar ao Mapa E é apresentado na Figura 3. A partir de um grupo de ligação com 19 marcadores e tamanho de 199.25 cM, foram obtidos dois grupos menores com 3.82 cM e 1.51 cM, os quais fazem parte do Mapa E. Neste caso, deixamos de utilizar 12 marcadores, no entanto os grupos menores foram formados utilizando critérios rigorosos como LOD score > 9.0 e fração de recombinação < 0.10 e representam porções confiáveis da estrutura genética da cana-de-açúcar. Em adição, estes grupos menores poderão compor um grupo maior com bom ordenamento quando mais marcadores forem amostrados na população de mapeamento. Desta forma, a utilização de um mapa com grupos de ligação menores, mas com bom ordenamento e densidade de marcadores, como o Mapa E, eleva a possibilidade de detecção de QTLs também confiáveis, como observado no Capítulo 2.

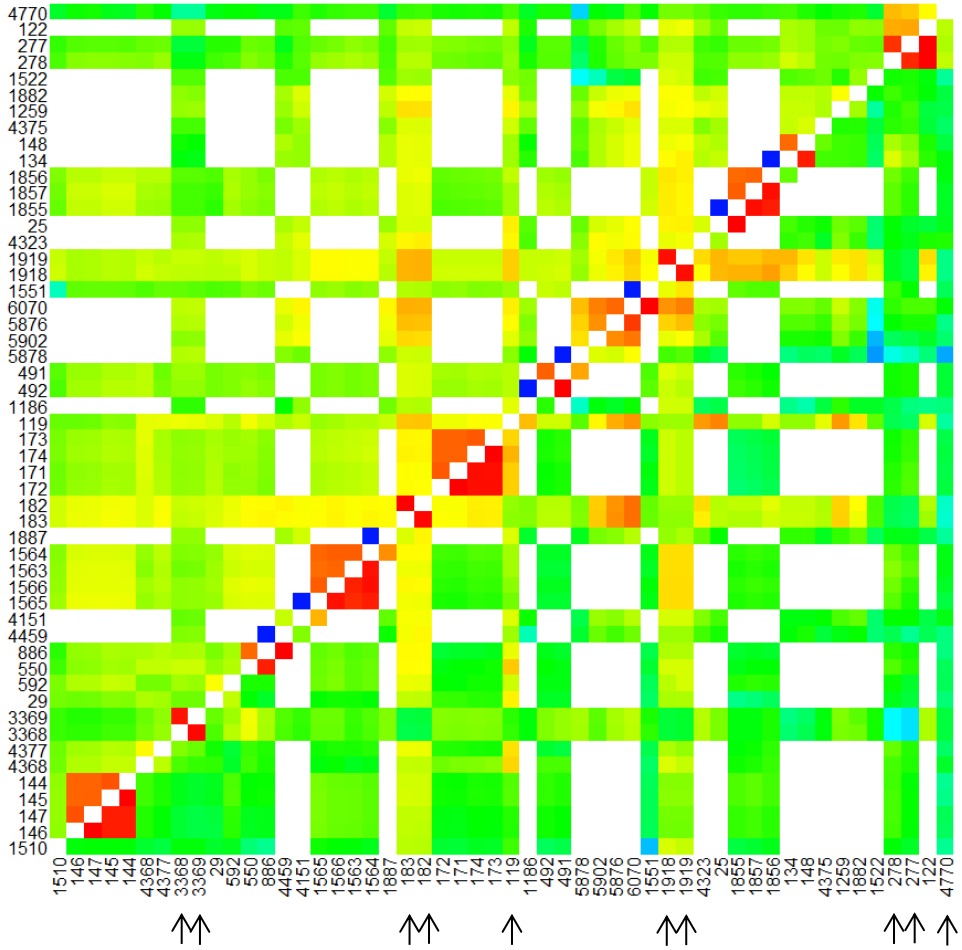


Figura 2 - *Heatmap* de um grupo de ligação do Mapa A formado quando utilizados LOD score > 9.0 e fração de recombinação < 0.15. Este grupo de ligação contou com 52 marcadores e apresentou tamanho total de 517.74 centimorgans (cM). Do total de marcadores, 10 e 42 são da classe de segregação 1:2:1 (indicados pela seta) e da classe de segregação 1:1 (25 marcadores D1.10, 13 marcadores D2.15 e 4 marcadores D2.18), respectivamente.

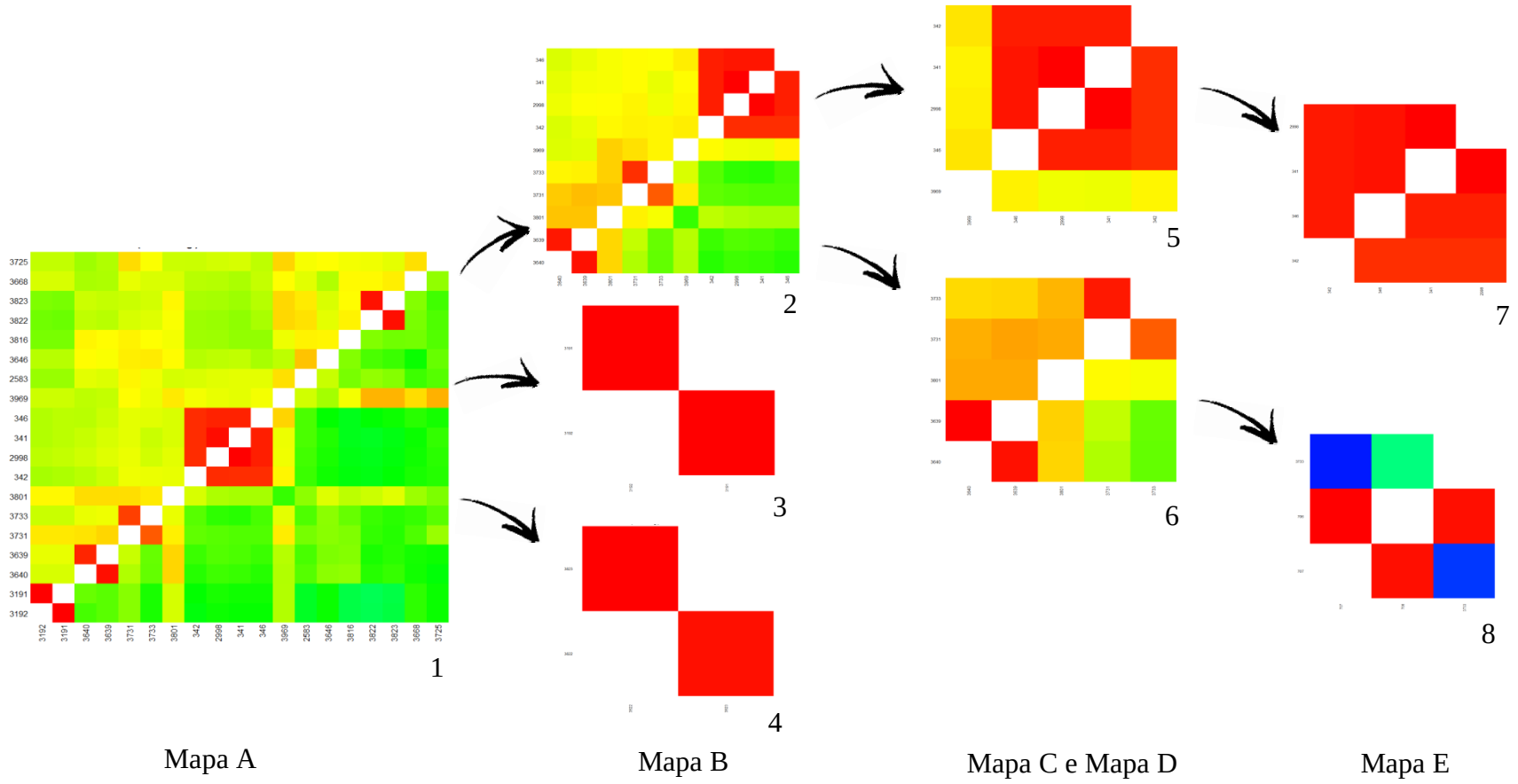


Figura 3 - *Heatmaps* exemplificando a obtenção do melhor ordenamento dentro de um grupo de ligação a partir do Mapa A até chegar ao Mapa E. O grupo de ligação obtido com o Mapa A (1) tem tamanho de 199.25 centimorgans (cM) e foi quebrado em outros três grupos menores (2, 3 e 4), os quais fazem parte do Mapa B. Os dois grupos menores que 1 centimorgan (cM) (3 e 4) foram excluídos e o grupo de tamanho maior, com 60.77 cM (2), foi quebrado em outros dois grupos com 16.23 cM (5) e 29.04 cM (6). Estes dois grupos fizeram parte dos Mapas C e D, visto que foram maiores que 1 cM e apresentaram mais que dois marcadores. O grupo 5 perdeu um marcador para formar o grupo 7 com 3.82 cM, enquanto que o grupo 6 perdeu dois marcadores para formar o grupo 8 com 1.51 cM.

Discussão geral

Aumentar a produtividade da cana-de-açúcar é uma árdua tarefa para agricultores, melhoristas e pesquisadores das áreas correlatas à cadeia produtiva. Além de dificuldades de diversas origens e magnitudes que o setor sucroenergético enfrenta constantemente (como por exemplo, crises financeiras internas e oscilações de preço do açúcar no mercado internacional) o pouco entendimento sobre a organização genética da cana-de-açúcar é um complicador a mais e que tem reflexo direto sobre o avanço de tecnologias moleculares importantes para esta cultura. Diferentemente do que ocorre para milho e soja, por exemplo, a cana-de-açúcar ainda caminha vagorasamente para utilizar marcadores moleculares no processo de seleção de novas cultivares. Essa lentidão não pode ser considerada como desinteresse de pesquisadores ou pela inexistência de pessoas qualificadas, pelo contrário, muitos grupos de pesquisa em diversas partes do mundo têm dedicado esforços para buscar o entendimento necessário sobre a complexidade genética da cana-de-açúcar e desenvolver novas ferramentas de geração e de análise de dados que almejam também a melhoria dos rendimentos agrícolas. No entanto, os investimentos para alavancar as pesquisas nesta área ainda estão abaixo do esperado frente à elevada complexidade do genoma da cana.

Diante de um cenário de estabilização dos ganhos genéticos relatada por alguns trabalhos, a utilização de ferramentas mais refinadas de análise de dados fenotípicos pode contribuir para obtenção de melhores estimativas e consequentemente melhores tomadas de decisões. Um programa de melhoramento genético convencional de cana-de-açúcar pode levar 15 anos para liberar comercialmente um novo cultivar. Todo o processo de seleção é de alto custo e envolve avaliações fenotípicas de diversas características em milhares de clones sendo que os melhores clones ainda são avaliados em vários locais. Assim, é fundamental que a metodologia de obtenção dos dados fenotípicos seja feita de forma padronizada e consistente, e que toda a análise desses dados seja realizada para considerar às complexas relações entre genótipos e ambientes. Em adição, experimentos realizados a campo devem ser conduzidos com rigor desde as fases iniciais do melhoramento, com utilização de padrões para comparação e delineamento estatístico adequado, a fim de garantir que as estimativas genéticas reflitam o real potencial de cada clone avaliado.

Os modelos mistos para análise de dados fenotípicos podem assumir heterogeneidade de variâncias genéticas e considerar dados perdidos, dados desbalanceados e acessar correlações genéticas entre ambientes, o que os torna apropriados para obtenção das estimativas dos valores genéticos de melhoramento, diretamente relacionados com

herdabilidade e ganho genético, em comparação com a simples média aritmética a qual foi e ainda é comumente praticada por programas de melhoramento para ranqueamento de clones durante as etapas de seleção. Demonstramos que as matrizes de efeitos genéticos ($\boldsymbol{G_p}$) podem ser diferentes para as características fenotípicas avaliadas e que também são diferentes quando comparadas famílias oriundas de *backgrounds* genéticos distintos. Em adição, verificamos que essas diferenças também podem acontecer para as matrizes de efeitos residuais ($\boldsymbol{R}$). Estes resultados são indicativos da complexa relação entre características e os ambientes. Os modelos mistos lineares para análise dos componentes de produção e o modelo misto linear generalizado para análise dos dados de severidade à ferrugem marrom foram eficientes na estimativa dos parâmetros genéticos e são ferramentas úteis na investigação de heterogeneidade de variâncias e correlações genéticas entre ambientes. Aliando experimentos a campo bem conduzidos com análise de dados capaz de obter importante informação acerca dos parâmetros genéticos, é possível orientar melhor os cruzamentos e selecionar os indivíduos mais produtivos de uma família durante as etapas de seleção de um programa de melhoramento. Os valores genéticos (*best linear unbiased estimation*, BLUP), dos indivíduos de uma família ainda podem ser utilizados por programas de mapeamento genético para identificação de regiões genômicas que controlam cacterísticas de interesse comercial em cana-de-açúcar.

A construção de mapas genéticos e o mapeamento de QTLs em organismos poliploides, como a cana-de-açúcar, são importantes para aumentar o conhecimento acerca da organização do genoma e abrir caminho para a utilização da seleção assistida por marcadores. Mapas genéticos com alta densidade de marcadores e alta cobertura do genoma são bons para atingir os objetivos anteriormente citados. No entanto, a complexidade do genoma da cana-de-açúcar (tamanho de aproximadamente 10 Gb, poliploide e com a ocorrência de aneuploidia), o custo para gerar grande quantidade de marcadores mapeáveis e a ausência de modelos genético-estatísticos capazes de considerar segregações de doses múltiplas tem limitado o desenvolvimento de mapas genéticos de alta densidade para esta cultura. Apesar dos esforços, ainda não temos publicado, até a data de redação desta tese, um genoma de referência completamente montado para cana-de-açúcar, fator que também dificulta a obtenção de mapas genéticos mais saturados, dentre outras atividades.

O uso de técnicas baseadas em sequenciamento de nova geração, como o GBS, pode contribuir para obtenção de maior quantidade de marcadores úteis na construção de mapas genéticos. Demonstramos que o GBS tem grande potencial para identificação de marcadores moleculares, mesmo com o emprego de pseudo-referências, os quais podem ter a

ploidia e dosagem alélica estimadas através do software SUPERMASSA e serem utilizados para construção de um mapa genético integrado de cana-de-açúcar juntamente com marcadores SSR e TRAP. Este mapa foi apto para o mapeamento de QTLs para quatro características fenotípicas importantes em cana-de-açúcar e que haviam sido avaliadas anteriormente através de modelo misto linear: conteúdo de sólidos solúveis (BRIX), conteúdo de sacarose na cana (POL%C), conteúdo de fibra (FIB) e diâmetro de colmos (SD). Para cana-de-açúcar, é importante que um QTL seja mantido entre diferentes locais e ambientes. Assim, a região do genoma envolvida com a característica pode ser estudada com o intuito de contribuir para o desenvolvimento de cultivares ecléticas, ou seja, que apresentam alta produtividade em condições edafoclimáticas diversas. Nossos resultados sugerem que BRIX e POL%C têm QTLs estáveis entre os dois locais avaliados, Araras e Ipaussu. Em adição, como uma primeira abordagem, as sequências que originaram os marcadores que flanquearam as regiões de QTL passaram por análise de homologia para verificação e predição de possíveis genes candidatos. Apesar do número de genes nas regiões de QTL ser incerto, BRIX e FIB podem ter potenciais marcadores para iniciar estudos de seleção assistida em cana-de-açúcar. No entanto, é necessário que pesquisas mais detalhadas sejam conduzidas para determinar quais e quantos genes das regiões de QTL estão envolvidos com as características e quais os reais efeitos sobre o fenótipo.

Resumo dos resultados

Os resultados obtidos e expostos a seguir de forma resumida demonstram que os objetivos inicialmente propostos foram alcançados.

Capítulo 1

- A abordagem de modelos mistos com a seleção de matrizes de variância-covariância, que avaliaram a heterogeneidade de variância genética e correlação genética entre ambientes para as características fenotípicas POL%C, POL%J, BRIX, TPH, TCH, SW, SN, SD, SH e FIB nas duas famílias de cana-de-açúcar (SR1 e SR2), permitiu obter estimativas eficientes dos parâmetros genéticos. Ao todo foram quatro os modelos selecionados: FA1, UNST, UNST ⊗ AR1 e UNST ⊗ UNST.

- Na seleção dos modelos para a matriz de efeitos genéticos ($\boldsymbol{G_p}$) em SR1, as características SD, SW, BRIX, POJ%J, FIB, TCH e TPH tiveram os parâmetros genéticos estimados considerando estruturas de variâncias e covariâncias genéticas para local e colheita separadamente, ao passo que para as características SH, SN e POL%C, as estruturas de variâncias e covariâncias genéticas possuíam uma combinação fatorial local-colheita. Já em SR2, com exceção para a característica SW, todas as demais tiveram os parâmetros genéticos estimados considerando estruturas de variâncias e covariâncias genéticas que combinaram os fatores local e colheita.

- Em geral, as herdabilidades em nível de média foram altas, variando de 0.78 (SH) a 0.92 (SD) em SR1 e de 0.79 (POL%C) a 0.94 (SD) em SR2.

- Foram detectadas 17 e 12 correlações genotípicas significativas entre as características avaliadas para SR1 e SR2, respectivamente.

- A análise dos dados de reação à ferrugem marrom via modelo misto linear generalizado revelou que 66% e 32% dos clones em SR1 e SR2, respectivamente, possuem, no mínimo, 90% de probabilidade de serem resistentes a esta doença.

Capítulo 2

- Utilizando a progênie do cruzamento entre as cultivares SP80-3280 e RB835486 (denominada de SR1 no capítulo anterior), o protocolo GBS forneceu, após as etapas de avaliação de cobertura e filtragens, mais de 99 mil marcadores considerando quatro pseudo-referências para a cana-de-açúcar: genoma de sorgo (*Sorghum bicolor*), genoma metil-filtrado da cana-de-açúcar, transcriptoma da cana-de-açúcar (RNAseq) e sequências do projeto SUCEST.

- Estes marcadores foram levados para estimação de ploidia e dosagem alélica no software SUPERMASSA. Dos mais de 99 mil marcadores, 12,551 foram classificados como de dose única. Os marcadores em dose única foram avaliados para redundância, restando 7,049 loci com chamada única.

- Os 7,049 marcadores oriundos do protocolo GBS foram adicionados com 629 marcadores em dose única baseados em gel (SSR e TRAP) para construção de um mapa genético integrado.

- O mapa final obtido possui 993 marcadores mapeados distribuídos ao longo de 223 grupos de ligação, 18 grupos de homo(eo)logia e com cobertura total de 3,682.04 cM. Com a estimativa do nível de ploidia e da dosagem alélica, foi possível incluir no mapa genético integrado marcadores com segregação 1:2:1 oriundos do GBS.

- Utilizando mapeamento por intervalo composto, procedemos a análise de mapeamento de QTL para quatro (POL%C, BRIX, FIB e SD) das 11 características fenotípicas avaliadas no Capítulo 1. Foram encontrados sete QTLs, sendo três para BRIX, dois para POL%C, um para FIB e um para SD.

- Os resultados da análise de mapeamento sugerem a presença de um QTL estável entre locais para conteúdo de sólidos solúveis (BRIX) e para teor de sacarose (POL%C). Além disso, QTLs para BRIX e teor de fibra (FIB) tiveram marcadores associados com genes candidatos com grande potencial de validação.

Conclusões

A avaliação de características fenotípicas através da abordagem de modelos mistos permitiu encontrar estimativas dos parâmetros genéticos com eficiência. As duas famílias de cana-de-açúcar avaliadas no Capítulo 1 apresentaram resultados distintos para as estruturas de variâncias e covariâncias genéticas selecionadas para cada característica, refletindo a complexidade das interações entre os ambientes e as características. Os modelos mistos utilizados para estimar os parâmetros genéticos podem ser testados em progênies de cana-de-açúcar de programas de melhoramento a fim de aumentar a eficiência dos processos de seleção. Além disso, o modelo misto linear generalizado utilizado na análise dos dados de severidade à ferrugem marrom pode ser utilizado pelos programas de melhoramento a fim de classificar os clones em níveis de probabilidade de suscetibilidade e/ou resistência.

A utilização de marcadores oriundos do protocolo de GBS, juntamente com marcadores baseados em gel, permitiu a obtenção de um mapa genético integrado com uma elevada densidade de marcadores e mostrou grande potencial para mapeamento de QTLs em cana-de-açúcar. A verificação e predição de genes candidatos para os QTLs mapeados mostrou ser importante para novas linhas de raciocínio sobre as complexas relações entre fenótipo e genótipo. Estudos mais detalhados são necessários para estabelecer os reais efeitos dos genes envolvidos com o fenótipo. Em adição, ainda são necessárias estratégias e ferramentas estatísticas capazes de incluir marcadores em múltiplas doses nos mapas genéticos de poliploides com o intuito de ligar os marcadores em dose única que estão "pulverizados" pelos vários cromossomos e promover eficiente mapeamento de genes que não são amostrados pela cobertura insuficiente do genoma.

Perspectivas

Esta tese contribui para aumentar o conhecimento sobre análises fenotípicas de características importantes em cana-de-açúcar e também sobre a utilização de marcadores identificados a partir de dados de GBS para construção de mapa genético e realização do mapeamento de QTLs. Com os dados obtidos nesta tese ainda deveremos trabalhar para a validação dos genes candidatos preditos que estão na região dos QTLs identificados. Assim, poderemos estudar de foma mais detalhada as possíveis vias metabólicas e redes regulatórias que estão ativas, e relacioná-las com o aumento ou diminuição da produtividade. Além disso, poderemos identificar potenciais marcadores para estudos de seleção assistida.

O estudo genético da cana-de-açúcar é um constante desafio visto que ainda caminhamos para encontrar as melhores metodologias de obtenção e de análises de dados genômicos e fenotípicos. Nosso grupo de pesquisa tem dedicado esforços para contribuir com o entendimento do complexo genoma da cana-de-açúcar. O desenvolvimento de técnicas da biologia molecular, como o GBS, e de ferramentas estatísticas, como o software SUPERMASSA, promovem cada vez mais a possibilidade de identificarmos regiões genômicas controladoras de características fenotípicas importantes. A associação de formas alélicas nas suas diferentes dosagens com níveis de expressão fenotípica e, também, a seleção genômica são objetivos almejados para o futuro e que poderão contribuir sobremaneira para aumentar a produtividade dos canaviais.

Referências

Aitken KS, McNeil MD, Hermann S, Bundock PC, Kilian A, Heller-Uszynska K, Henry RJ, Li J: A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. *BMC Genomics*, 2014 a. 15:152.

Aitken KS, Hermann S, Karno K, Bonnett GD, McIntyre CL, Jackson PA: Genetic control of yield related stalk traits in sugarcane. *Theoretical and Applied Genetics*, 2008. 117:1191-1203.

Aitken KS, Jackson PA, McIntyre CL: Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum spp*.) cultivar x *Saccharum officinarum* population. *Theoretical and Applied Genetics*, 2006. 112:1306-1317.

Al-Janabi SM, Parmessur Y, Kross H, Dhayan S, Saumtally S, Ramdoyal K, Autrey LJC, Dookun-Saumtally A: Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. *Molecular Breeding*, 2007. 19:1-14.

Alwala S, Suman A, Arro JA, Veremis JC, Kimbeng CA: Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science*, 2006. 46:448-455.

Asnaghi C, D'hont A, Glaszmann JC, Rott P: Resistance of sugarcane cultivar R570 to *Puccinia melanocephala* from different geographic locations. *Plant Dis*, 2001. 85:282–286.

Batley J, Barker G, O´Sullivan H, Edwards KJ, Edwards D: Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant physiology*, 2003. 132(1):84-91.

Empresa de Pesquisa Energética (EPE), Ministério de Minas e Energia (MME). Balanço energético nacional 2015: Ano base 2014. Rio de Janeiro, 2015. 292 p.

Buetow KH, Edmonson MN, Cassidy AB: Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genetics*, 1999. 21(3) 323-325.

Bundock PC, Henry RJ: Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. TAG. *Theoretical and applied genetics*, 2004. 109(3):543-551.

Bundock, PC, Eliott FG, Ablett G, et al.: Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant biotechnology journal*, 2009. 7(4)347–54.

Byrne S, Czaban A, Studer B, Panitz F, Bendixen C, Asp T: Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. *PloS One*, 2013. *8*(3), e57438.

Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, Carneiro MS, Garcia AAF, de Souza AP, Vicentini R: De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One*, 2014. 9:e88462.

Cheavegatti-Gianotto A, Abreu HMC De, Arruda P, et al.: Sugarcane (*Saccharum officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Tropical plant biology*, 2011. 4(1):62–89.

Chen AH, Lipka AE: The Use of Targeted Marker Subsets to Account for Population Structure and Relatedness in Genome-Wide Association Studies of Maize (*Zea mays* L.). G3, 2016. doi: 10.1534/g3.116.029090.

Cho RJ, Mindrinos M, Richards DR, et al.: Genome-wide mapping with biallelic markers in Arabidopsis thaliana. *Nature genetics*, 1999. v. 23, n. 2, p. 203-207, 1999.

CONAB Companhia Nacional de Abastecimento. Disponível em: http://www.conab.gov.br/OlalaCMS/uploads/arquivos. Acessado em 03 de outubro de 2016.

Cordeiro GM, Eliott F, McIntyre CL, Casu RE, Henry RJ: Characterisation of single nucleotide polymorphisms in sugarcane ESTs. *Theoretical and applied genetics*, 2006. 113(2):331-343.

Cordeiro GM, Pan Y-B, Henry RJ: Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. *Plant Science*, 2003. 165(1):181–189.

Costa EA, Anoni CO, Mancini MC, Santos FRC, Marconi TG, Gazaffi R, Pastina MM, Perecin D, Mollinari M, Xavier MA, Pinto LR, Souza AP, Garcia AAF: QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny. *Euphytica*, 2016. 221(1):1-16.

Creste S, Sansoli DM, Tardiani ACS, Silva DN, Goncalves FK, Favero TM, Medeiros CNF, Festucci CS, Carlini-Garcia LA, Landell MGA, Pinto LR: Comparison of AFLP, TRAP and SSRs in the estimation of genetic relationships in sugarcane. *Sugar Tech*, 2010. 12:150-154.

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, Babu R: Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3*, 2013. *3*(11):1903–26.

D'Hont A and Glaszmann JC: Sugarcane genome analysis with molecular markers, a first decade of research. *Proc. Int. Soc. Sugarcane. Technol.*, 2001. 24: 556-559.

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC: Determination of basic chromosome numbers in the genus Saccharum by physical mapping of ribosomal RNA genes. *Genome*, 1998. 41:221-225

D'Hont A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet Genome Res*., 2005. 109:27-33.

D'Hont A, Grivet L, Feldmann P, Rao P, Berding N, Glaszmann JC: Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. *Mol Gen Genet*, 1996. 250:405–413.

Dal-Bianco M, Carneiro MS, Hotta CT et al.: Sugarcane improvement: how far can we go? *Current opinion in biotechnology*, 2012. 23(2):265-70.

Daniels, J., Roach, B.T. (1987). Taxonomy and evolution. Chapter 2. In: DJ Heinz, ed. Sugarcane improvement through breeding. *Elsevier*, 1987. 11:7-84.

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, D'Hont A: A putative major gene for rust resistance linked with a RFLP marker in sugar cane cultivar 'R 570'. *Theoretical and Applied Genetics*, 1996. 92:1059-1064.

Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG: Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PloS One*, 2013. *8*(5), e62137.

Edwards MD, Stuber CW, Wendel JF: Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics, 1987. 116:113-125.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 2011. 6:e19379.

Falconer DS, Mackay TFC: Introduction to Quantitative Genetics. 4.ed. Essex, UK: Longman, 1996. 464 p.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). 2015. FAOSTAT. [Database] FAO, Rome. Disponível em: http://faostat3.fao.org/home/E. Acessado em: 03 de outubro de 2016.

Ferreira ME, Grattapaglia D: Introdução ao uso de marcadores moleculares em análise genética. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.

Garcia AAF, Mollinari M, Marconi TG, Serang O, Silva RR, Vieira MLC, Vicentini R, Costa EA, Mancini M, Garcia MO, Pastina MM, et al.: SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci Rep*. 2013. 3:3399. doi: 10.1038/srep03399.

Gazaffi R, Margarido GRA, Pastina MM, Mollinari M, Garcia AAF. A model for quantitative trait loci mapping, linkage phase and segregation pattern estimation for a full-sib progeny. *Tree Genet Genomes*. 2014. 10:791-801.

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, Souza AP: Development of na integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum likelihood approach for estimation of linkage and linkage phases. *Theor Appl Genet*, 2006. 112:298-314.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q et al.: TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*, 2014. 9: e90346.

Grivet L, Glaszmann J-C, Vincentz M, Da Silva F, Arruda P: ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. TAG. *Theoretical and applied genetics*, 2003. 106(2):190-197.

Guimarães CT, Sills GR, Sobral BWS: Comparativemapping of Andropogoneae: Saccharum L. (sugarcane) and its relation to sorghum and maize. *Proceedings of the National Academy of Science of the United States of America*, 1997. 94:14261-14266.

Henry RJ, Edwards M, Waters DLE, et al.: Application of large-scale sequencing to marker discovery in plants. *Journal of Biosciences*, 2012. 37(5):829–841.

Heslot N, Rutkoski J, Poland J, Jannink J-L, Sorrells ME: Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. *PLoS One*, 2013. 9:e74612. doi:10.1371/journal.pone.0074612.

Hoarau JY, Grivet L, Offmann B, Raboin LM, Diorflar JP, Payet J, Hellmann M, D'hont A, Glaszmann JC: Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.) II. Detection of QTLs for yield components. *Theor Appl Genet*, 2002. 105:1027-1037.
Jansen RC, Stam P: High resoluction of quantitative traits into multiple loci via interval mapping. *Genetics*, 1994. 136:1447-1455.

Jiang C, Zeng ZB: Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics, 1995. 140:1111-1127.

Jiang C, Zeng Z-B: Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica*, 1997, 101:47-58.

Jordan DR, Casu RE, Besse P, Carroll BC, Berding N, McIntyre CL: Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. *Genome*, 2004. 47:988-993.

Kanazin V, Talbert H, See D, Decamp P, Nevo E, Blake T: Discovery and assay of single-nucleotide polymorphisms in barley (Hordeum vulgare). *Plant molecular biology*, 2002. 48(5-6):529-537.

Kao CH, Zeng ZB, Teasdale R: Multiple interval mapping for quantitative trait loci. *Genetics*, 1999. 152:1203-1216.

Kao CH, Zeng ZB: General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics*, 1997. 53:653-665.

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A: Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Molecular genetics and genomics*, 2003. 270(1):24-33.

Lander ES, Botstein D: Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 1989. 121:185-199.

Lin M, Lou XY, Chang M, Wu RL. A general statistical framework for mapping quantitative trait loci in nonmodel systems: Issue for characterizing linkage phases. *Genetics.* 2003. 165:901–913.

Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, Waugh R: An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. *BMC Genomics*, 2014. 15(1), 104.

Liu BH: Statistical genomics: linkage, mapping, and QTL analysis. Boca Raton, USA: CRC Press. 1998. 611p.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Costich DE: Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, 2013. *9*(1), e1003215.

Lynch M, Walsh B: Genetics and analysis of quantitative traits. Sunderland, Massachusetts, USA: Sinuauer Associets, Inc. 1998. 980p.

Malosetti M, Ribaut JM, Vargas M, Crossa J, Eeuwijk FA van: A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica*, 2008. 241-257.

Malosetti M, Voltas J, Romagosa I, Ullrich SE, Eeuwijk FA van. Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica*, 2004.137:139-145.

Marconi TG, Costa EA, Miranda H, Mancini MC, Cardoso-Silva CB, Oliveira KM, Pinto LR, Mollinari M, Garcia A, Sousa AP: Functional markers for gene mapping and genetic diversity studies in sugarcane. *BMC Research Notes*, 2011. 4:264.

Margarido, GRA: Mapeamento de QTLs em múltiplos caracteres e ambientes em um cruzamento comercial de cana-de-açúcar usando modelos mistos. 2011. 107 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

Margarido GRA, Pastina MM, Souza AP, Garcia AAF: Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. *Molecular Breeding*, 2015. 35(8):175.

Mascher M, Wu S, Amand PS, Stein N, Poland J: Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PloS One*, 2013. 8(10), e76925.

McIntyre CL, Whan VA, Croft B, Magarey R, Smith GR: Identification and Validation of Molecular Markers Associated with Pachymetra Root Rot and Brown Rust Resistance in Sugarcane Using Map- and Association-based Approaches. *Molecular Breeding*, 2005. 16(2):151–161.

McIntyre CL, Casu RE, Drenth J, Knight D, Wham VA, Croft BJ, Jordan DR: Manners JM: Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance. *Genome*, 2005b. 48:391-400.

McIntyre CL, Jackson M, Cordeiro GM, Amouyal O, Hermann S, Aitken KS, Eliott F, Henry RJ, Casu RE, Bonnett GD: The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. *Molecular Breeding*, 2006. 18: 39-50.

McIntyre CL, Whan VA, Croft B, Magarey R, Smith GR: Identification and validation of molecular markers associated with Pachymetra root rot and brown rust resitance in sugarcane using map- and association-based approaches. *Molecular Breeding*, 2005a. 16:151-161.

Ming R, Liu SC; Moore PH, Irvine JE, Paterson AH: QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. *Genome Research*, 2001. 11:2075-2084.

Ming R, Delmonte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH: Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. *Genome,* 2002a. 45:794-803.

Ming R, Wang YW, Draye X, Moore PH, Irvine JE, Paterson AH: Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. *Theoretical and Applied Genetics*, 2002b. 105:332-345.

Mogg R, Batley J, Hanley S, Edwards D, O'sullivan H, Edwards J: Characterization of the flanking regions of Zea mays microsatellites reveals a large number of useful sequence polymorphisms. Theoretical and applied genetics., v. 105, n. 4, p. 532-543, 2002.

Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, Souza AP: Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Sacharum* spp.) commercial cross. *Molecular Breeding*, 2007. 20:189-208.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF and Souza AP: Characterization of new polymorphic functional markers for sugarcane. *Genome*, 2009. 52:191-209.

Palhares AC, Rodrigues-Morais TB, Sluys MV, Domingues DS, Junior WM, Junior HJ, Souza AP, Marconi TG, Mollinari M, Gazaffi R, Garcia AAF, Vieira MLC: A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC Genetics,* 2012. 13:51.

Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GRA, Oliveira KM, Pinto LR, Souza AP, Eeuwijk FA van, Garcia AAF: A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theoretical and Applied Genetics*, 2012. 124:835-849.

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, Souza AP de: Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. *Plant Breed* 2006, 125:378-384.

Pinto LR, Garcia AAF, Pastina MM, Teixeira LHM, Bressiani JA, Ulian EC, Bidoia MAP, Souza AP: Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. *Euphytica*, 2010. 172:313-327.

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, Souza AP: Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. 2004. 804:795-804.

Piperidis N, Jackson PA, D'hont A, Besse P, Hoarau JY, Courtois B, Aitken KS, Mcintyre CL: Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. *Molecular Breeding*, 2008. 21:233-247.

Piperidis G, Piperidis N, D'Hont A: Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics*, 2010. 284:65-73.

Poland JA, Brown PJ, Sorrells ME, Jannink JL: Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS One*, 2012 a. *7*(2). doi:10.1371/journal.pone.0032253.

Poland JA, Rife TW: Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome Journal*, 2012 b. 5(3):92. doi:10.3835/plantgenome2012.05.0005.

Raboin L-M, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau J-Y, D'Hont A: Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theor Appl Genet* 2006, 112:1382-91.

Raboin LM, Pauquet J, Butterfield M, D'hont A, Glaszmann JC: Analysis of genome-wide linkage disequilibriumin the highly polyploid sugarcane. *Theoretical and Applied Genetics*, 2008. 116:701-714.

Rafalski A, Tingey S: SNPs and their use in maize. Plant genotyping II: SNP TECHNOLOGY. Oxfordshire, UK: CAB International, 2008. p. 30-43.

Reffay N, Jackson PA, Aitken KS, Hoarau JY, D'hont A, Besse P, McIntyre CL: Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding*, 2005. 15:367-381.

Robasky K, Lewis NE, Church GM: The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics*, 2014. *15*(1), 56–62.

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Gardner CA: Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, 2013. *14*(6), R55.

Serang O, Mollinari M, Garcia AAF: Efficient Exact Maximum a Posteriori Computation for Bayesian SNP Genotyping in Polyploids. *PlosOne,* 2012. 7(2):e30906. doi:10.1371/journal.pone.0030906. 2012.

Sills GR, Bridges W, Al-Janabi SM, Sobral BWS: Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. ex Grassl). *Molecular Breeding*, 1995. 1:355-363.

Singh R, Mishra S, Singh S, Mishra N, Sharma M: Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. *Aust. J. Crop Sci*, 2010. 4(2):115–124.

Smith AB, Cullis BR, Thompson R: Analyzing variety by environment data using multiplicativemixed models and adjustments for spatial field trend. *Biometrics*, 2001. 57:1138-1147.

Sonah H, Bastien M, Iquira E, Tardivel A, Legare G, Boyle B, et al.: An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE*, 2013. 8:e54603. doi: 10.1371/journal.pone.0054603

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Mccouch S: Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theoretical and Applied Genetics*, 2013. *126*(11), 2699–716. doi:10.1007/s00122-013-2166-x.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS: Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). *Proceedings of the National Academy of Sciences of the United States of America*. 2001. 98(16):9161-9166.

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA.; Visser RGF, Van Eck HJ: A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PloS One*, 2013. *8*(5), e62355. doi:10.1371/journal.pone.0062355.

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti E a, Lemos MVF, Coutinho LL, Nobrega MP, Carrer H, França SC, Bacci Júnior M, Goldman MHS, Gomes SL, Nunes LR, Camargo LE a, Siqueira WJ, Van Sluys M-A, Thiemann OH, Kuramae EE, Santelli R V, Marino CL, Targon MLPN, Ferro J a, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, et al.: Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res*, 2003. 13:2725-35.

Wei X, Jackson PA, McIntyre CL, Aitken KS, Croft B: Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics*, 2006. 114:155-164.

Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR: A comparison of analysis methods for late-stage variety evaluation trials. *Australian & New Zealand Journal of Statistics*, 2010. 52:125-149.

Weller JI: Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics*, 1986. 42:627-640.

Yu J, Hu S, Wang J, et al.: A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 2002. 296:79-92.

Zeng ZB: Statistical methods for mapping quantitative trait loci. Raleigh: Department of Statistics, North Carolina State University, 2001. 128 p.

Zeng ZB: Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proceedings of the National Academy of Science of the United States of America*, 1993. 90:10972-10976.

Zeng ZB: Precision mapping of quantitative trait loci. *Genetics*, 1994. 136(4):492-496.

Zeng ZB, Kao CH, Basten CJ: Estimating the genetic architecture of quantitative traits. *Genetical Research*, 1999. 74:279-289.

Anexo I

OPEN ACCESS Freely available online

PLOS ONE

De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties

Claudio Benicio Cardoso-Silva[1☯], Estela Araujo Costa[1☯], Melina Cristina Mancini[1], Thiago Willian Almeida Balsalobre[1], Lucas Eduardo Costa Canesin[1], Luciana Rossini Pinto[2], Monalisa Sampaio Carneiro[3], Antonio Augusto Franco Garcia[4], Anete Pereira de Souza[1,5], Renato Vicentini[1]*

1 Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil, **2** Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana (IAC/Apta), Ribeirão Preto, SP, Brazil, **3** Departamento de Biotecnologia e Produção Vegetal e Animal, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras, SP, Brazil, **4** Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, Brazil, **5** Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

Abstract

Sugarcane is an important crop and a major source of sugar and alcohol. In this study, we performed *de novo* assembly and transcriptome annotation for six sugarcane genotypes involved in bi-parental crosses. The *de novo* assembly of the sugarcane transcriptome was performed using short reads generated using the Illumina RNA-Seq platform. We produced more than 400 million reads, which were assembled into 72,269 unigenes. Based on a similarity search, the unigenes showed significant similarity to more than 28,788 sorghum proteins, including a set of 5,272 unigenes that are not present in the public sugarcane EST databases; many of these unigenes are likely putative undescribed sugarcane genes. From this collection of unigenes, a large number of molecular markers were identified, including 5,106 simple sequence repeats (SSRs) and 708,125 single-nucleotide polymorphisms (SNPs). This new dataset will be a useful resource for future genetic and genomic studies in this species.

Citation: Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, et al. (2014) *De Novo* Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. PLoS ONE 9(2): e88462. doi:10.1371/journal.pone.0088462

Editor: Cynthia Gibas, University of North Carolina at Charlotte, United States of America

Received August 15, 2013; **Accepted** January 7, 2014; **Published** February 11, 2014

Funding: The authors gratefully acknowledge the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, http://www.fapesp.br) for the financial support grants 2008/52197-4 (AS) and 2008/58031-0 (RV) and for the graduate scholarships to CBCS, EAC, MCM, and TWB, and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ, http://www.cnpq.br) for the research fellowships to AAG, APS, and RV. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shinapes@unicamp.br

☯ These authors contributed equally to this work.

Background

Sugarcane belongs to the grass family (Poaceae), which is an economically important seed plant family that includes maize, wheat, rice, sorghum and many types of grasses. The sugarcane crop is the main source of both sugar and alcohol, accounting for two-thirds of the world's sugar production [1]. It is estimated that approximately 653.81 million tons of sugarcane will be produced during the 2013/2014 harvest in Brazil, surpassing the production of the last harvest [2].

Modern sugarcane varieties are derived from interspecific hybridization between *Saccharum officinarum* and *Saccharum spontaneum*, resulting in highly polyploid and aneuploid plants. Indeed, the chromosome number of these varieties ranges from 80 to 140. Modern varieties of sugarcane typically exhibit more than eight homologous copies of each basic chromosome from *S. officinarum* and several copies of the homologous chromosomes from *S. spontaneum* [3]. Therefore, sugarcane cultivars are highly heterozygous, presenting several different alleles at each locus, and this high level of genetic complexity creates challenges during conventional and molecular breeding programs.

Recent technological developments have the potential to greatly increase our understanding of sugarcane plants through the application of emerging genomic technologies, and the use of next-generation sequencing (NGS) technologies could have significant implications for crop genetics and breeding. Although the sequencing of large genomes remains expensive, even using NGS technologies [4], transcriptome sequencing can provide information regarding the gene content of a species and can complement genome sequencing approaches.

RNA sequencing (RNA-Seq) has been applied as a tool for transcriptome analysis in many species, such as *Arabidopsis thaliana* [5], *Brassica* spp. [6], rice [7] and maize [8]. RNA-Seq has several advantages, including (i) allowing more precise measurement of the levels of transcripts and their isoforms than other methods, (ii) presenting the potential for the development of SNPs that can be used to detect allele-specific expression because the same base is sequenced multiple times, (iii) the ability to identify reads containing post-transcriptional modifications or rearranged sequences that cannot be mapped directly to the genome [9] and (iv) allowing the identification of species-specific genes [10]. Moreover, the availability of a large number of genetic markers developed using NGS technologies is facilitating trait mapping and marker-assisted breeding [11].

In plant breeding programs, genotypes of interest to breeders, such as the parental genotypes of mapping populations, can be

sequenced using NGS technologies. More than one genotype can be employed to generate sequence data with these technologies, and these data can be aligned using genome or transcriptome sequencing data for model or major crop species that are closely related to the species of interest [11]. This approach has also been applied for marker discovery in some crop species, such as eucalyptus [12], maize [13] and chickpea [14], and has been used to identify SNPs between the parental genotypes of mapping populations. These SNPs can then be employed to develop markers for marker-deficient crops to allow trait mapping through marker-assisted selection (MAS).

Despite its economic importance, no published genome sequence is currently available for sugarcane. Instead, the basic resource used for the study of sugarcane gene sequences is the substantial expressed sequence tag (EST) information available in public databases. Transcriptome studies in sugarcane began in South Africa [15,16], and the largest EST collection (~238,000 ESTs) was developed through the Brazilian SUCEST project [17,18]. Researchers in Australia [19–21] and the USA [22] have generated three additional libraries containing 10,000 ESTs each. Currently, all of the reported ESTs are collected in the Sugarcane Gene Index, version 3.0, which contains 282,683 ESTs and 499 complete cDNA sequences, resulting in 121,342 unique assembled sequences, or unigenes. There are still more than 10,000 sugarcane coding genes that have yet to be identified [23], highlighting the need for new sequencing efforts in the sugarcane transcriptome. This information would increase the panel of potential molecular markers and sequence information available for sugarcane breeding programs, resulting in biotechnological improvements. In the present study, using the Illumina GA IIx sequencing platform, we performed *de novo* transcriptome sequencing in six sugarcane genotypes that are employed as parents in Brazilian Sugarcane Breeding Programs. We identified conserved genes that have not previously been described in sugarcane, and these data will be useful for future genome assembly and marker identification.

Materials and Methods

Ethics Statement

We confirm that no specific permits were required for the described field studies. This work was a collaborative research project developed by researchers from UNICAMP, ESALQ/USP, IAC/Apta (Instituto Agronômico de Campinas) and UFSCar-RIDESA (Universidade Federal de São Carlos-Rede Interinstitucional de Desenvolvimento do Setor Sucroalcooleiro) (all from Brazil). We also confirm that the field studies did not involve endangered or protected species.

Plant Materials and RNA Extraction

Six genotypes were included in this study. IACSP96-3046 and IACSP95-3018 are the parents of a mapping population from the Sugarcane Breeding Program at IAC/Apta. IACSP95-3018 is a promising clone that is also used as a parent in the breeding program. IACSP93-3046 is a variety that exhibits good tillering, an erect stool habit [24] and resistance to rust [25].

SP81-3250×RB925345 and SP80-3280×RB835486 are the parents of two different mapping populations from the Sugarcane Breeding Program at UFSCar, which is part of RIDESA. These parents exhibit contrasting properties: SP81-3250 and SP80-3280 are resistant to rust [26,27], whereas RB925345 and RB835486 are susceptible [28]. All of the examined genotypes display high levels of sucrose.

Leaves at the third position [29] were collected from one plant per genotype and immediately frozen, and total RNA was extracted using a modified protocol [30]. The integrity and quantity of the isolated RNA were assessed using a 2100 Bioanalyzer (Agilent). Equal quantities of high-quality RNA from each genotype were pooled for cDNA synthesis.

mRNA-Seq Library Construction for Illumina Sequencing

Paired-end Illumina mRNA libraries were generated from 4 µg of total RNA in accordance with the manufacturer's instructions for mRNA-Seq Sample Preparation (Illumina Inc., San Diego, CA, USA). The quality of the library was assessed using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

Cluster amplification was performed using the TruSeq PE Cluster Kit and a cBot (Illumina), and each sample was sequenced in a separate GAIIx lane using the TruSeq SBS 36 Cycle Kit (Illumina). The read length was 72 bp.

Sequence Data Analysis and Assembly

The raw data generated by Illumina sequencing were converted from the BCL format to qSeq using Off-line Basecaller, v.1.9.4 (OLB) software. The qSeq files were transformed in FastQ files, which contain sequences that are 72 bp in length, using a custom script. Low-quality sequences were removed; these sequences included reads with ambiguous bases, reads with less than 70 bases, and reads with a Phred quality score Q≤20 using the NGS QC toolkit [31]. All reads were deposited in the National Center for Biotechnology Information (NCBI) database and can be found under accession number SRA073690.

All datasets were combined, and the sequenced reads were assembled using Trinity (http://trinityrnaseq.sourceforge.net/), which is a program developed specifically for *de novo* transcriptome assembly from short-read RNA-Seq data that recovers transcript isoforms efficiently and sensitively using the de Bruijn graph algorithm [32]. The optimal assembly results were chosen according to an evaluation of the assembly encompassing the total number of contigs, the distribution of contig lengths, the N50 statistic and the average coverage. The assembled transcripts were based on the main isoform of each transcript, and only contigs with lengths of greater than 300 bp were included in the downstream analysis.

To identify the genotypic contribution to each transcript, reads from each library were mapped against the assembly generated from all libraries using the bowtie aligner [33]. The BAM files generated by bowtie were then used to estimate the transcript-level abundance for each library using the RSEM (RNA-Seq by Expectation Maximization) software [34].

Functional Annotation of Sugarcane Transcripts

The assembled sequences were compared against the NCBI non-redundant protein database (NR) using BLASTX with a cut-off E-value of 10^{-6}. To annotate the assembled sequences according to Gene Ontology (GO) terms (The Gene Ontology Consortium, 2000), the above BLAST results were analyzed using Blast2GO [35] to determine and compare gene functions. The GO terms were assigned to the representative transcripts for each sample through an enrichment analysis using Fisher's exact test (p-value <0.01), with a false discovery rate (FDR) correction in terms of biological processes and molecular functions. The transcript sequences were also aligned against the *Viridiplantae*, grass and sorghum protein databases (http://www.phytozome.org/) using BLASTX and against the Sugarcane Gene Index (http://compbio.dfci.harvard.edu/tgi/) using BLASTN; in both alignments, a cut-off E-value of 10^{-6} was applied. The BLAST search

was limited to the first ten significant query hits, and the gene names were assigned to each query based on the highest score. Transcripts that showed similarity to *Viridiplantae* proteins were aligned against the sorghum genome using sim4 software [36]. Open reading frames (ORFs) were predicted using a script available in the TransDecoder package (http://transdecoder.sourceforge.net/), with 300 bp as the minimum ORF length. Those transcripts showing predicted ORFs were aligned against grass proteins using the STRING database, v.9.05 (http://string-db.org), to predict Clusters of Orthologous Groups (COG).

To further characterize the subset of unigenes that did not show similarity to any known plant proteins, we applied a computational strategy to mine putative long non-coding RNA (lncRNA) data. We first aligned all 121,342 EST unigenes to *Viridiplantae* proteins and to the GenBank NR database using BLASTX. Those EST unigenes that did not align with any proteins were then mapped to the *Sorghum bicolor* genome, obtaining at least 70% coverage and a maximum intron size of 15 kb. The coding probability of the positively mapped unigenes was then evaluated by removing sequences with potential ORFs longer than 100 aa using ESTScan [37]. We further investigated the functional role of the remaining unigenes and putative lncRNAs by searching for three indirect indications of functionality: we examined the stability of the secondary structure using the Vienna package [38], normalized to the Z-score index [39]; we mapped the small RNAs (sRNAs) [40] against sugarcane unigenes; and we analyzed the sequence similarities between the unigenes and *S. bicolor* ESTs (BLASTN, E-value $\leq 1e^{-5}$). Only EST unigenes with at least one indirect piece of functional evidence were analyzed further. The putative lncRNAs were then aligned to the 18,910 assembled transcripts that showed no similarity to any plant protein but were successfully mapped to *S. bicolor* (Text S4). Only hits with an E-value below $1e^{-5}$ and coverage higher than 40% were considered positive.

Putative Molecular Markers

We utilized the MISA program (http://pgrc.ipk-gatersleben.de/misa/) to search for simple sequence repeat (SSR) motifs in the unigenes; the MISA script can identify both perfect and compound (interrupted by a certain number of bases) motifs. To identify the presence of SSRs, only motifs of two to six nucleotides were considered, and the minimum repeat unit was defined as six for dinucleotide motifs and five for tri-, tetra-, penta- and hexanucleotide motifs. A compound motif was defined as two or more SSR motifs interrupted by sequences of up to 100 bp.

To identify putative single-nucleotide polymorphisms (SNPs) in the sugarcane transcript assembly, we first separately mapped all of the short reads from each library to the assembly using the Burrows-Wheeler Aligner (BWA). Next, FreeBayes [41] and SAMtools [42] were used to detect the variable positions of SNPs from the consensus sugarcane assembly. The FreeBayes tool allowed us to identify genetic variants in the polyploid organisms. The putative SNPs were then filtered using the varFilter command, where variants were called only for positions with a minimal mapping quality (-Q) and coverage (-d) of 25. To compare the composition of the SNP variation in the parental genotype, unique and shared SNPs were extracted using an in-house script. The transition and transversion ratios were calculated using the tstv tool developed by SnpSift software [43].

Results and Discussion

De novo assembly of the sugarcane transcriptome

The libraries sequenced using the Illumina platform produced a total of 610,232,490 paired-end (PE) sequence reads, each of which was 72 bp in length. We filtered the sequence data for low-quality reads, resulting in 445,374,504 high-quality PE trimmed reads (97.67%), which were used to obtain the *de novo* assembly. An overview of the sequencing procedure is presented in Table 1. The *de novo* assembly generated 119,768 transcripts when all isoforms were considered. These transcripts represent a total of 72,269 unigenes that were considered for downstream analysis (Text S1). The length of the unigenes ranged from 300 bp to ~7 kb, with a mean length of 921 bp, an N50 equal to 1,367 bp and 46.39% GC content. The average length of the assembled unigenes was greater than those obtained from chickpea (523 bp) [14], rubber trees (485 bp) [44] and bamboo (736 bp) [45] using similar sequencing technologies. Considering the N50 values, the values for the sugarcane unigenes were greater than those for rubber trees (592 bp), bamboo (1,132 bp) and chili pepper (1,076 bp) [46], which were also assembled using short reads generated by the Illumina platform. In total, we obtained 18,624 (27.21%) unigenes longer than 1 kb and 7,657 (10.6%) unigenes longer than 2 kb. The length distributions of the unigenes are shown in Table 2, revealing that more than 40,000 unigenes (55.76%) were longer than 500 bp. These unigenes were submitted to an ORF predictor using TransDecoder, and we detected 33,673 (46.59%) unigenes with ORFs, with 9,350 (12.94%) presenting complete ORFs.

Unigene annotation

The 72,269 sugarcane unigenes were analyzed for sequence similarity against the *Viridiplantae* (comprising all green plants) and grass (*S. bicolor*, *Oryza sativa*, *Zea mays*, *Panicum virgatum*, *Setaria italica* and *Brachypodium virgatum*) datasets through BLASTX searches. The unigenes were also compared against the sugarcane EST database via a BLASTN search (Table 3). A total of 35,456 (49.06%) unigenes showed significant similarity to *Viridiplantae*. The high percentage of sugarcane unigenes obtained in this study that did not match the *Viridiplantae* protein database (50.84%) indicates that there is potential for the discovery of as-yet-undescribed and novel genes in sugarcane, although most of these unigenes may encode non-coding RNAs. In fact, more than 26% of the unigenes in this set exhibited high similarity to intergenic regions of the sorghum genome (Figure 1). Additionally, the significance of a BLAST search depends on the length of the query sequence; therefore, short sequences are rarely matched to known genes [12], or these sequences may represent rapidly evolving sequences that have diverged substantially from their homologs [47].

In turn, alignment of the unigenes against the grass protein database returned 34,814 significant hits. When considering the hits by species, 28,788 unigenes showed significant similarity to sorghum, corresponding to 98% of sorghum proteins (Figure 1).

Table 1. Summary of Illumina transcriptome sequencing data for the sugarcane varieties included in this study.

Sample	Read length (bp)	Raw data	Trimmed data	GC (%)	Q20 (%)
SP95-3018	72+72	84,105,462	64,906,391	49.04	98.09
SP81-3250	72+72	103,971,718	71,002,186	47.52	97.32
RB925345	72+72	112,124,334	77,476,268	46.91	97.11
SP80-3280	72+72	101,983,186	73,160,814	47.59	97.56
RB835486	72+72	119,280,444	87,873,521	46.62	97.66
SP93-3046	72+72	88,767,346	70,955,324	48.07	98.25

doi:10.1371/journal.pone.0088462.t001

Table 2. Summary of the *de novo* assembly results for the sugarcane transcriptome.

Unigene length (bp)	Total unigenes	Percentage
300–500	31,971	44.24%
500–1000	20,634	28.55%
1000–2000	12,007	16.61%
2000–3000	4,827	6.68%
3000–4000	1,790	2.47%
4000–5000	636	0.88%
>5000	404	0.56%
Total length (bp)	66,572,642	-
Unigenes	72,269	-
N50 length	1,367	-
GC (%)	46.39	-

doi:10.1371/journal.pone.0088462.t002

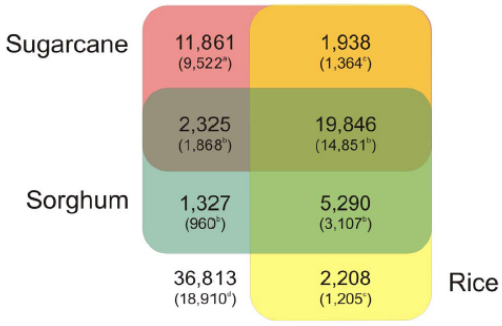


Figure 1. Proportions of sugarcane transcripts showing homology to sugarcane unigenes and sorghum and rice proteins. For annotation, the best BLASTX/N hit against the protein or nucleotide sequences of the reference organisms was employed, with an E-value cut-off of $\leq 10^{-6}$. The number between the parentheses indicates the number of different proteins/unigenes in each species (sugarcane[a], sorghum[b] and rice[c]). The number outside of the Venn diagram indicates no-hit transcripts and the number of transcripts[d] that mapped to the sorghum genome.
doi:10.1371/journal.pone.0088462.g001

These results were expected, as comparative genomic studies [48] have revealed conservation and synteny among the sugarcane and sorghum genomes. The sugarcane transcriptome also significantly matched that of rice, with approximately 29,285 unigenes (corresponding to 28,732 unique protein accessions) showing significant similarity to rice proteins.

To investigate previously unidentified potential genes in sugarcane, we compared the unigenes against the sugarcane transcripts deposited in public databases and performed BLAST searches to detect possible similarities with the SoGI database (*S. officinarum*). Furthermore, the unigenes that did not show similarity to sugarcane ESTs were compared against sorghum proteins. Approximately 22,171 unigenes exhibited significant similarity to sorghum proteins and sugarcane transcripts (Figure 1). The remaining 5,272 unigenes (Text S3) showed significant similarity to sorghum and rice proteins but not to the sugarcane transcripts that were considered to be putative new sugarcane genes (Figure 1). By examining the presence of candidate coding regions in these unigenes, we identified 4,895 sequences that contained ORFs, with 732 unigenes containing complete ORFs. These unigenes represent genes that have not yet been described for sugarcane.

Clusters of Orthologous Groups (COG) classification

COG classification was performed for the transcriptome data, and a total of 7,519 unigenes were identified (Figure 2). These unigenes were classified into 23 COG categories, with the largest number of unigenes being grouped in the 'replication, recombination and repair' cluster (20.49%), followed by the 'general function prediction only' cluster (17.05%) and the 'posttranslational modification, protein turnover and chaperones' cluster (7.39%). These three categories are the same categories that are highly represented in sorghum (Figure 2).

A total of 19 of the 23 COG categories were present in the transcriptome data, and at least 60% of the sugarcane unigenes were annotated when compared with the annotation of sorghum genes in the COG categories.

The categories 'energy production and conversion' (3.72%), 'carbohydrate transport and metabolism' (5%) and 'defense mechanisms' (2%) exhibited at least 56% of the expected genes compared with the sorghum genes. These categories should be considered to represent gene sequences showing a high potential for the development of molecular markers in sugarcane breeding programs. Therefore, the likelihood of these markers being associated with agronomic traits of interest in QTL mapping and marker-assisted selection (MAS) [49] is increased.

Table 3. Summary of the annotation of each database.

Database	Number of unigenes	Number of proteins matched	Percentage of unigenes[a]
Viridiplantae proteins	35,456	34,969	49.06%
Grass proteins	34,814	34,304	48.17%
Sorghum proteins	28,788	28,030	39.83%
Hits against sorghum proteins and sugarcane ESTs	22,171	20,969	30.68%
Total of no-hit unigenes	36,813	-	50.94%
No-hit unigenes with high similarity to the sorghum genome	18,910	-	26,16

[a]Percentage relative to the total number of sugarcane unigenes.
doi:10.1371/journal.pone.0088462.t003

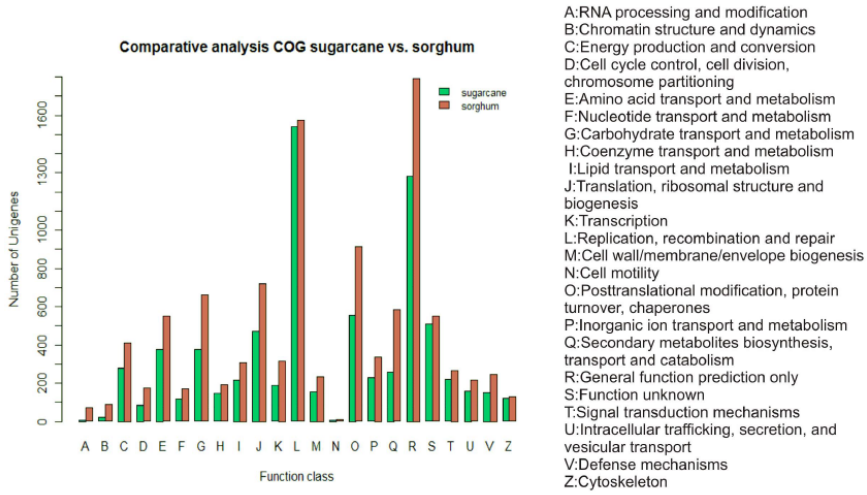


Figure 2. Histogram of the Clusters of Orthologous Groups (COG) classifications of the sugarcane transcripts and sorghum proteins.
doi:10.1371/journal.pone.0088462.g002

Gene Ontology enrichment analyses

The identification of functional classes that differ statistically between two lists of terms is a typical data-mining approach applied in functional genomics research [35]. In this work, we were interested in identifying which functions were distinctly represented among the different sugarcane genotypes. A total of 14,983 unigenes (Text S2) were annotated based on BLAST matches to known proteins in the NR database and were assigned to GO classes representing 39 terms, including some (10) that contain important information related to the enriched genotype (Figure 3).

Genes responsible for disease resistance, corresponding to the categories 'signaling,' 'response to stimulus,' 'cellular response to stimulus,' 'response to chemical stimulus' and 'response to auxin stimulus', were enriched in the SP81-3250, SP80-3280 and IACSP93-3046 genotypes, with IACSP93-3046 being represented in all of these categories (Figure 3). These three genotypes exhibit resistance to rust [25–27], whereas the other genotypes, RB925345, RB835486 and IACSP95-3018, are susceptible to rust [24,28]. Common sugarcane rust, caused by the fungus *Puccinia melanocephala*, is a disease that occurs worldwide and can result in large losses of sugar tonnage in susceptible varieties [50]. Rust resistance is generally considered to be a quantitatively inherited trait showing a high degree of heritability and a strong additive genetic variance component [51,52].

The obtained enriched terms suggest that these three genotypes harbor transcripts that are involved in stimulus response pathways and probable disease responses. These results are correlated with the characteristics of resistance and susceptibility in these varieties.

Another important characteristic of sugarcane crops is their accumulation of sucrose. Wild sugarcane species produce less than 4% fresh weight of sucrose, whereas high-yield varieties can produce sucrose contents of up to 20% of their fresh weight [53]. The major differences between these varieties is based on sugar transport and metabolism in storage tissues [54]. The entire network involving sucrose synthesis, accumulation, storage and retention is a complex system in which several metabolic pathways interact with each other [55]. The most important aspect of this network is transport, which chiefly involves specific carrier molecules, ion transport and active transport and depends on the amount of available ATP. Within this context, we observed some genotypes that were enriched in categories related to this network, particularly the transport process. These categories included 'organic substance transport' (SP81-3250, RB925345, SP80-3280, IACSP96-3046 and IACSP95-3018), 'substrate-specific transporter activity', 'substrate-specific transmembrane transporter activity' (SP81-3250 and SP80-3280), 'ion transmembrane transport' (SP81-3250 and IACSP93-3046) and 'transporter activity' (SP81-3250, SP80-3280, and IACSP93-3046).

Important categories involved in sugar transport and metabolism in storage tissues include the 'monosaccharide metabolic process,' 'glucose metabolic process,' 'small molecule biosynthetic process' and 'small molecule metabolic process' categories. The terms in the first and second categories were only enriched in the SP81-3250 genotype, whereas the terms in the third category were enriched in both the IACSP93-3046 and IACSP95-3018 genotypes. All genotypes showed enrichment in the last category, although SP80-3280 was the least represented.

All of the genotypes were enriched for transcripts involved in this complex network of sucrose synthesis, accumulation, storage and retention, and these results were corroborated by the agronomic characteristics of the plants. All of these genotypes produce high levels of sucrose, in accordance with the agronomic description of the genotypes SP81-3250 [26], RB925345, RB835486 [28], SP80-3280 [27], IACSP93-3046 [25] and IACSP95-3018 [24].

Putative lncRNAs

Among the initial set of 121,342 EST retrieved unigenes, 23,529 showed no similarity to any known plant protein. These unigenes were mapped to the *S. bicolor* genome, resulting in 4,476 positive hits, with only 1,884 not exhibiting an ORF or presenting an ORF

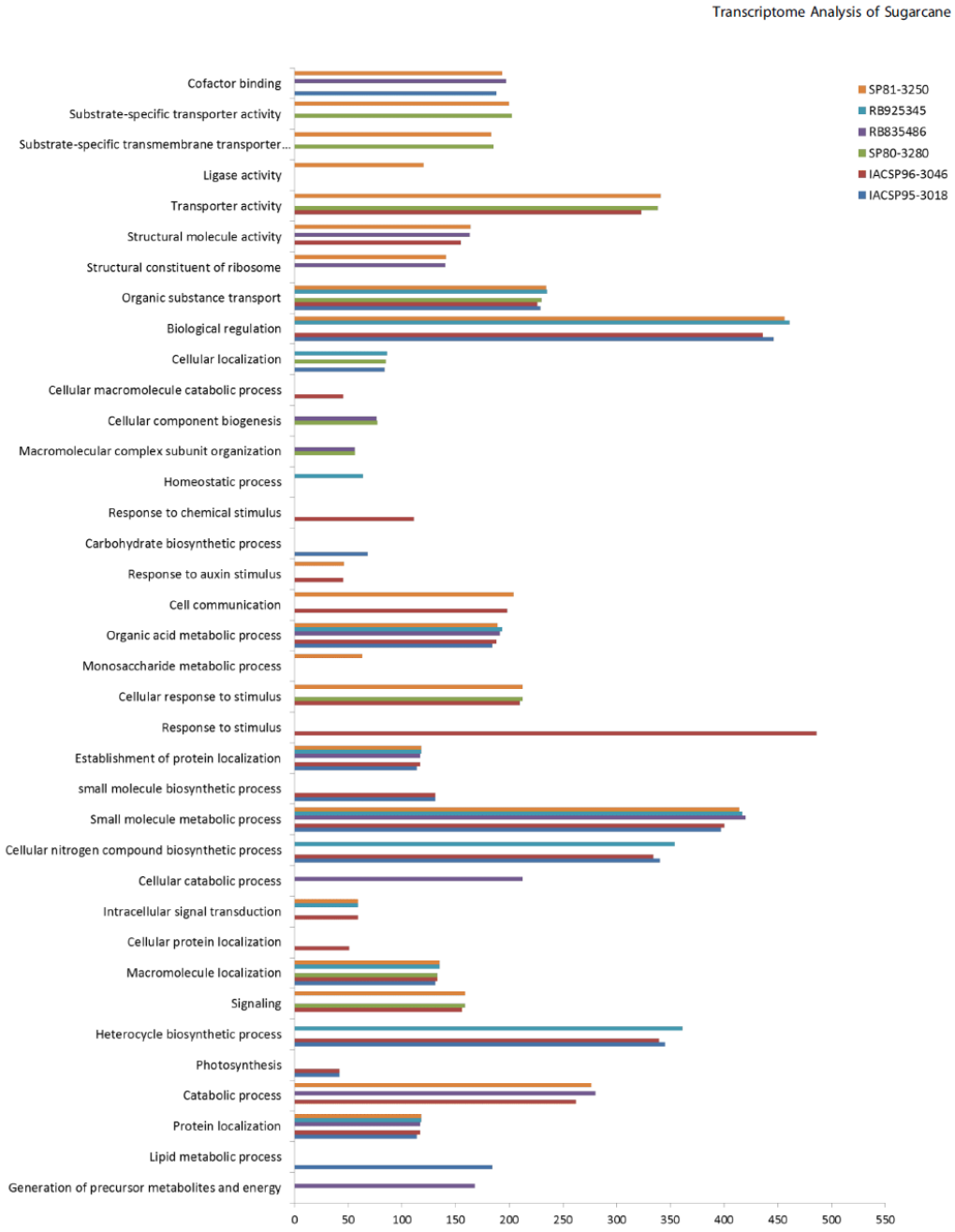


Figure 3. Enrichment of Gene Ontology terms for each sugarcane variety.
doi:10.1371/journal.pone.0088462.g003

shorter than 100 aa. This subset comprised the putative sugarcane lncRNAs that are publicly available. We found that for ~4% of these sequence, there were small RNAs (sRNAs) that mapped to their sequence, with ~59% showing similarity to *S. bicolor* and ~39% showing a highly stable secondary structure. In total, 1,446 non-redundant putative lncRNAs were identified that showed indirect evidence of functionality (Figure S1). We then compared this inclusive set (1,884 sequences) with the 18,910 assembled transcripts that lacked similarity to plant proteins. We observed 358 putative lncRNAs represented among the assembled transcripts, with ~42% of these sequences showing a highly stable secondary structure and ~40% showing evidence of transcription in the *S. bicolor* EST dataset. None of the unigenes to which sRNAs were mapped were similar to any assembled transcript. Finally, we compared the expression profiles of the putative lncRNAs between the different genotypes, which suggested that these transcripts may display genotype-specific expression patterns, as shown in Figure 4. A hierarchical clustering analysis revealed a pattern of separation between the genotypes from the different breeding programs, a result that is in accordance with the observation that the varieties from the same breeding program have the same genetic basis. We observed that the plant lncRNAs may display elevated intraspecific variation in expression, and several recent works have demonstrated that these transcripts exhibit tissue- and cell-specific expression patterns [56–59]. This study adds information regarding the dynamic involvement of these transcripts and reveals putative targets for further investigation [60,61].

Marker discovery

SSR discovery. Expressed sequence tag/simple sequence repeat (EST-SSR) markers are well established as important tools for researchers assessing genetic diversity and are useful in the development of genetic maps, comparative genomics and MAS breeding. Thus, the unigene sequences were searched for repeat motifs to explore the SSR profiles in the sugarcane transcriptome. A total of 5,106 SSRs were obtained from 4,616 unigene sequences (7.96%), and 576 of the unigenes contained more than one SSR (Text S7). Of these unigenes, 189 exhibited compound SSR formation. Trinucleotide repeat motifs were the most abundant, accounting for 2,585 SSRs (50.63%) in 2,318 unigene sequences; dinucleotide repeat motifs accounted for 1,927 SSRs (37.74%) in 1,732 unigenes; and other motifs accounted for 594 SSRs (11.63%) in 1,708 unigenes (Table 4). The relative percentage of the sequences containing SSRs was higher than that obtained in the SUCEST (Sugarcane Expressed Sequence Tag database) study, in which 2,005 clusters containing SSRs were found among 43,141 clusters (4.64%) [62].

The most abundant motifs included the dinucleotide AG motif (49.9%) and the trinucleotide CCG (17%) and ACC (4.7%) motifs. These results are similar to those of the SSR motif analysis performed in sorghum [63]. Additionally, CCG and ACC were the most commonly found motifs in the SUCEST study [62], and CCG was the motif that was identified most often by Cordeiro *et al.* [64]. The most frequent tetranucleotide motif found in the

Table 4. Summary of the simple sequence repeat (SSR) types in the sugarcane transcriptome.

Repeat motif	Number[a]	Unigenes[b]	Percentage (%)[c]
Di-nucleotide			
AC/GT	551		
AG/CT	962		
AT/TA	336		
CG/GC	78		
Total	**1,927**	**1,732**	**37.74**
Tri-nucleotide			
AAC/GTT	141		
AAG/CTT	152		
AAT/ATT	60		
AGC/GCT	219		
ACG/CGT	197		
AGT/ACT	62		
ACC/GGT	122		
AGG/CCT	252		
ACA/TGT	97		
AGA/TCT	46		
ATA/TAT	24		
ATC/GAT	42		
ATG/CAT	43		
CAC/GTG	69		
CAG/CTG	228		
CCG/CGG	442		
CGC/GCG	241		
CTC/GAG	148		
Total	**2,585**	**2,318**	**50.63**
Other motifs[d]	**594**	**1,708**	**11.63%**
Total	**5,106**	**5,758**	-

[a]Number of the total SSRs (di-, tri- and other motifs).
[b]Number of unigene sequences containing SSRs.
[c]The relative percentage of SSRs with different repeat motifs among the total SSRs.
[d]The total number of SSRs of other sizes.
doi:10.1371/journal.pone.0088462.t004

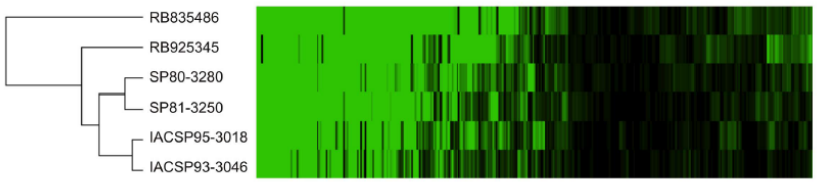


Figure 4. Hierarchical clustering of the 358 putative sugarcane lncRNAs. The expression patterns allowed the identification of the genotypes based on their ability to store sucrose and according to the bi-parental crosses involved in the different mapping populations.
doi:10.1371/journal.pone.0088462.g004

present study was AAAG. The overall frequency of SSRs was observed to be 1/1.6 kb.

The prevalence of trimeric motifs over other SSR repeats may be explained based on the risk of frameshift mutations that may occur when microsatellites alternate in size [65]. Furthermore, a large number of trinucleotide coding repeats appear to be controlled primarily by mutation pressure.

The development of SSR markers associated with important agronomic traits can be used to assist in the selection of varieties during the early stages of MAS breeding programs and can be helpful in the selection of the best parents for crossing [66]. Consequently, the application of such markers supports breeding programs by significantly reducing the time and cost involved in developing new varieties and can help bypass barriers in sugarcane breeding programs.

SNP discovery. A total of 708,125 putative SNP positions were identified (Text S5), with a density of 1 SNP per 86 bp. The frequency of SNPs found in the sugarcane genes was higher than has been observed in other grasses, such as rice and sorghum, which exhibit a frequency of ≥1 SNP per 300 bp [67]. The observed number of transitions was 456,666, and 254,658 transversions were detected, with the number of the former being 1.79 times that of the latter. Transitions were most likely more frequent because they are more tolerated by natural selection as the tendency to generate synonymous mutations in coding sequences is related to the number of transversions [68].

We identified SNPs in 58,903 different unigenes, which represent 81.50% of the total unigenes. Considering the number of unigenes without SNPs, we verified that 10,516 (79%) are unigenes with a length of less than 500 bp. Considering only those unigenes with predicted ORFs (33,673 unigenes), we found a total of 289,969 SNPs (37.5% of the total detected SNPs).

To detect different heterozygous SNPs between the parents from each mapping population, the reads from each genotype were mapped against all the unigenes (Text S6). Figure 5 shows the heterozygous SNPs that were detected, and the unique and shared SNPs in each parent from the mapping populations were evaluated. The percentages of SNPs that were common in the three mapping populations, IACSP95-3018×IACSP93-3046 (32.86%), SP81-3250×RB925345 (32.42%) and SP80-3280×RB835486 (34.06%), were similar, and these SNPs may thus be polymorphic between the parents. As sugarcane is a polyploid species, polymorphisms can be generated from a different number of allelic copies present in each genotype. However, such polymorphisms are difficult to validate (Garcia *et al* 2013, *submitted*).

The SNPs that were unique to each genotype (Figure 5) exhibited a higher probability of association with the contrasting agronomic traits of interest. Because polymorphism markers between parents are important for generating saturated genetic mapping in mapping populations, these SNPs are a source of data for generating markers associated with quantitative trait loci (QTLs). Such functional molecular markers have been broadly applied for the genetic improvement of several crops [69].

According to the Gene Ontology annotation, we identified SNPs in 6,712 unigenes with annotation information, representing 44.80% of the unigenes included in the enrichment analyses. Some categories exhibited important results related to the genotype (Figure 3), particularly those associated with disease resistance. In the 'signaling' category, we identified 161 unigene sequences with SNPs, whereas we identified 477 unigenes with SNPs in the 'response to stimulus' category. These unigenes likely represent source data for the development of functional markers related to disease resistance.

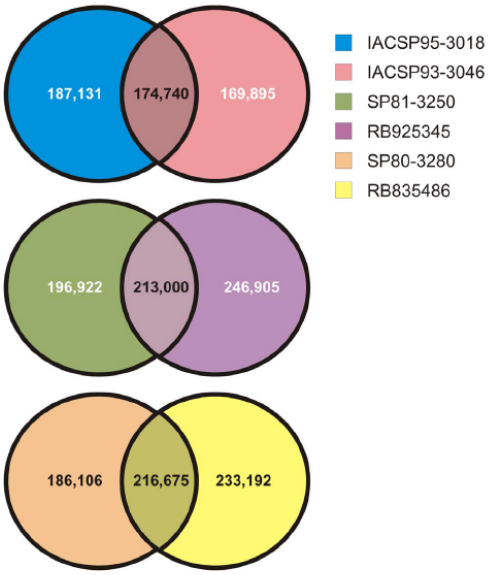


Figure 5. Unique and shared heterozygous putative SNPs in the parental genotypes of the three sugarcane mapping populations.
doi:10.1371/journal.pone.0088462.g005

When we analyzed the categories related to sucrose synthesis, accumulation, storage and retention, we also observed unigenes with SNPs in the 'organic substance transport' (226), 'substrate-specific transporter activity' (196) and 'ion transmembrane transport' (53) clusters. Equally important categories involving sugar transport and metabolism in storage tissues, such as the 'glucose metabolic process' (43), 'small molecule biosynthetic process' (133) and 'small molecule metabolic process' (414) categories, also containing unigene sequences with SNPs.

All of these unigene sequences with SNPs represent an important source of data. These sequences could be priority candidates for the development of specific functional markers and could be very useful in further genetic or genomic studies in sugarcane.

Conclusion

This is the first publicly available sugarcane transcriptome sequencing study performed using NGS technology to investigate the entire sugarcane transcriptome, and our data provide the most comprehensive transcriptome resource currently available for sugarcane. In addition, polymorphisms associated with candidate genes potentially involved in the stimulus response, energy production and growth were identified among the contrasting varieties and deserve future investigation. Based on the enrichment analysis, we identified putative genes related to disease and the accumulation of sucrose. Additionally, a large number of SNPs and SSRs were identified, and marker development would be a useful resource for future genetic or genomic studies of this species. Finally, this work contributed information on 5,000 undescribed

genes, which is more than half of the expected sugarcane genes that are missing from sugarcane databases.

Supporting Information

Figure S1 Venn diagram showing the classification of the identified putative sugarcane lncRNAs in the EST data (A) and RNA-Seq data (B).
(TIF)

Text S1 Unigene sequences in FASTA format.
(ZIP)

Text S2 Gene ontology enrichment annotation for the transcripts of each genotype.
(ZIP)

Text S3 Putative previously unknown sugarcane transcripts showing the best matches to sorghum proteins.
(TXT)

Text S4 List of 18,910 putative sugarcane ncRNAs with high coverage in the sorghum genome.
(TXT)

Text S5 List of 708,125 putative SNP positions identified in this study.
(ZIP)

Text S6 List of putative SNPs identified in each genotype.
(ZIP)

Text S7 List of 5,106 putative SSR positions identified in this study.
(XLS)

Author Contributions

Conceived and designed the experiments: AAFG MSC LRP APdS RV. Performed the experiments: EAC MCM TWAB. Analyzed the data: CBCS EAC LECC RV. Contributed reagents/materials/analysis tools: EAC MCM TWAB. Wrote the paper: CBCS EAC RV.

References

1. United States Department of Agriculture (2013) Sugar: World Markets and Trade. Foreign Agric Service. Available: http://usda01.library.cornell.edu/usda/current/sugar/sugar-11-21-2013.pdf. Accessed 10 December 2013.
2. Ministério da Agricultura (2013) Acompanhamento de safra brasileira: cana-de-açúcar Safra 2012/2013 Terceiro levantamento. Cia Nac Abast. Available: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_12_12_10_34_43_boletim_cana_portugues_12_2012.pdf. Accessed 10 December 2013.
3. Ming R, Liu SC, Lin YR, da Silva J, Wilson W, et al. (1998) Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150: 1663–1682.
4. Li S-W, Yang H, Liu Y-F, Liao Q-R, Du J, et al. (2012) Transcriptome and gene expression analysis of the rice leaf folder, Cnaphalocrosis medinalis. PLoS One 7: e47401.
5. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.
6. Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnol J 7: 334–346.
7. Lu T, Lu G, Fan D, Zhu C, Li W, et al. (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res 20: 1238–1249.
8. Hansey CN, Vaillancourt B, Sekhon RS, de Leon N, Kaeppler SM, et al. (2012) Maize (Zea mays L.) genome diversity as revealed by RNA-sequencing. PLoS One 7: e33071.
9. Marguerat S, Bähler J (2010) RNA-seq: from technology to biology. Cell Mol Life Sci 67: 569–579.
10. Morozova O, Marra Ma (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92: 255–264.
11. Varshney RK, Nayak SN, May GD, Jackson Sa (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27: 522–530.
12. Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, et al. (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genomics 9: 312.
13. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51: 910–918.
14. Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18: 53–63.
15. Carson DL, Botha FC (2000) Preliminary Analysis of Expressed Sequence Tags for Sugarcane. Crop Sci 40: 1769–1779.
16. Carson D, Botha F (2002) Genes expressed in sugarcane maturing internodal tissue. Plant Cell Rep 20: 1075–1081.
17. Vettore AL, Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol Biol 24: 1–7.
18. Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13: 2725–2735.
19. Casu RE, Grof CPL, Rae AL, McIntyre CL, Dimmock CM, et al. (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol 52: 371–386.
20. Casu RE, Dimmock CM, Chapman SC, Grof CPL, McIntyre CL, et al. (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol Biol 54: 503–517.
21. Bower NI, Casu RE, Maclean DJ, Reverter A, Chapman SC, et al. (2005) Transcriptional response of sugarcane roots to methyl jasmonate. Plant Sci 168: 761–772.
22. Ma H-M, Schulze S, Lee S, Yang M, Mirkov E, et al. (2004) An EST survey of the sugarcane transcriptome. Theor Appl Genet 108: 851–863.
23. Vicentini R, Bem LEV., Sluys Ma., Nogueira FTS, Vincentz M (2012) Gene Content Analysis of Sugarcane Public ESTs Reveals Thousands of Missing Coding-Genes and an Unexpected Pool of Grasses Conserved ncRNAs. Trop Plant Biol 5: 199–205.
24. Mancini MC, Leite DC, Perecin D, Bidóia MaP, Xavier Ma., et al. (2012) Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. Sugar Tech 14: 119–125.
25. Landell MGA, Campana MP, Figueiredo P, Vasconcelos ACM, Xavier MA, Bidoia MAP, Prado H, Silva MA, Miranda LLD AC (2005) Variedades de cana-de-açúcar para o centro sul do Brasil. Technical Bulletin IAC 197: 33.
26. Bellodi N, Macedo I (1995) Quinta geração de variedades de cana-de-açúcar. COOPERATIVA DOS PRODUTORES DE CANA, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO. Technical Bulletin: 16–23.
27. Sabino J (1997) Sexta geração de variedades de cana-de-açúcar. COOPERATIVA DE PRODUTORES DE CANA, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO LTDA. Technical Bulletin: 1.
28. Hoffmann H (2008) Variedades RB de cana-de-açúcar. CCA/UFSCar Technical Bulletin 1: 30.
29. McCormick AJ, Cramer MD, Watt DA (2006) Sink strength regulates photosynthesis in sugarcane. New Phytol 171: 759–770.
30. Kistner C, Matamoros M (2005) RNA ISOLATION USING PHASE EXTRACTION AND L I C L. In: Márquez A, editor. Lotus japonicus Handbook. Dordrecht, The Netherlands. pp. 123–124.
31. Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7: e30619.
32. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652.
33. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25–R25.
34. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323.
35. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.
36. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A Computer Program for Aligning a cDNA Sequence with a Genomic DNA Sequence. Genome Res 8: 967–974.
37. Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: A Program for Detecting, Evaluating, and Reconstructing Potential Coding Regions in EST Sequences. ISMB-99 Proceedings. AAAI Press. pp. 138–148.
38. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, et al. (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6: 26.
39. Clote P, Ferré F, Kranakis E, Krizanc D (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. RNA 11: 578–591.

40. Domingues DS, Cruz GMQ, Metcalfe CJ, Nogueira FTS, Vicentini R, et al. (2012) Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. BMC Genomics 13: 137.
41. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. Genomics (q-bioGN); Quant Methods: 1–9.
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
43. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, et al. (2012) Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 3: 35.
44. Li D, Deng Z, Qin B, Liu X, Men Z (2012) De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (Hevea brasiliensis Muell. Arg.). BMC Genomics 13: 192.
45. Liu M, Qiao G, Jiang J, Yang H, Xie L, et al. (2012) Transcriptome sequencing and de novo analysis for Ma bamboo (Dendrocalamus latiflorus Munro) using the Illumina platform. PLoS One 7: e46766.
46. Liu S, Li W, Wu Y, Chen C, Lei J (2013) De Novo Transcriptome Assembly in Chili Pepper (Capsicum frutescens) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. PLoS One 8: e48156.
47. Vincentz M, Cara FAA, Okura VK, da Silva FR, Pedrosa GL, et al. (2004) Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol 134: 951–959.
48. Grivet L, Hont AD, Dufour P, Hamon P, Roquest D (1994) Comparative genome mapping of sugar cane with other species within the Andropogoneae tribe. Heredity 73: 500–508.
49. Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3: 22–32.
50. Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, et al. (1996) A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570 '. Theor Appl Genet 92: 1059–1064.
51. Tai PYP, Miller JD, Dean JL (1981) INHERITANCE OF RESISTANCE TO RUST IN SUGARCANE. F Crop Res 4: 261–268.
52. Hogarth DM, Ryan CC, Taylor PWJ (1993) Quantitative inheritance of rust resistance in sugarcane. F Crop Res 34: 187–193.
53. Irvine JE (1975) Relations of Photosynthetic Rates and Leaf and Canopy Characters to Sugarcane Yield. Crop Sci 15: 671.
54. Moore PH, Botha F, Furbank R, Grof CP (1996) Intensive sugarcane production: Meeting the challenge beyond 2000. Keating BA and . Wilson JR, editor Oxon, UK: CAB International. p544.
55. Henry R, Kole C (2010) Genetics, Genomics and Breeding of Sugarcane. 1st ed. Henry, R. J,:. Kole C, editor Science Publishers. p300.
56. Guo X, Gao L, Liao Q, Xiao H, Ma X, et al. (2013) Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. Nucleic Acids Res 41: e35.
57. Hangauer MJ, Vaughn IW, McManus MT (2013) Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. PLoS Genet 9: e1003569.
58. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22: 1775–1789.
59. Liu J, Jung C, Xu J, Wang H, Deng S, et al. (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. Plant Cell 24: 4333–4345.
60. Sun J, Zhou M, Mao Z-T, Hao D-P, Wang Z-Z, et al. (2013) Systematic analysis of genomic organization and structure of long non-coding RNAs in the human genome. FEBS Lett 587: 976–982.
61. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, et al. (2013) Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLoS Genet 9: e1003470.
62. Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, de Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47: 795–804.
63. Ramu P, Kassahun B, Senthilvel S, Ashok Kumar C, Jayashree B, et al. (2009) Exploiting rice-sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet 119: 1193–1204.
64. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160: 1115–1123.
65. Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10: 72–80.
66. Marconi TG, Costa EA, Miranda HR, Mancini MC, Cardoso-Silva CB, et al. (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. BMC Res Notes 4: 264.
67. Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, et al. (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14: 1812–1819.
68. Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Tree 11: 158–162.
69. Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7: 132–136.

Anexo II

Crop Breeding and Applied Biotechnology 11: 280-285, 2011
Brazilian Society of Plant Breeding. Printed in Brazil

RB965902 and RB965917 – Early/medium maturing sugarcane varieties

Monalisa Sampaio Carneiro[1]*, João Ricardo Bachega Feijó Rosa[1], Fernanda Zatti Barreto[1], Thiago Willian Almeida Balsalobre[1], Roberto Giacomini Chapola[1], Marcos Antonio Sanches Vieira[1], Antonio Ismael Bassinello[1] and Hermann Paulo Hoffmann[1]

Received 21 September 2010

Accepted 24 February 2011

ABSTRACT – *The varieties RB965902 and RB965917 were developed for harvesting at the beginning to the middle of the sucrose extraction period (early/medium maturity) and released for the South-Central region of Brazil. In specific environments, the tons of Pol per area (sucrose yield) of these varieties is higher than of the commercial standard RB855453 and they are resistant to the main diseases of the crop.*

Key words: Saccharum, *Rideso*, *sugarcane breeding program.*

INTRODUCTION

The complex *Saccharum* spp. (known as sugarcane) is believed to be originated from complex natural hybridization events (called nobilization) between *Saccharum officinarum*, *S. barberi*, *S. sinense*, and related wild species *S. spontaneum* (Sreenivasan et al. 1987). Sugarcane is predominantly allogamous, highly heterozygous, and vegetatively propagated. There are currently four sugarcane breeding programs in Brazil: the Agronomical Institute of Campinas, whose varieties are labeled with the abbreviation IAC; the Center of Sugarcane Technology, with the identification code CTC, the Inter University Network for the Development of Sugar and Alcohol – RIDESA, with the varietal acronym RB varieties (Republic of Brazil), and CanaVialis/Monsanto with the initials CV.

RIDESA is a partnership of 10 Federal Universities with the purpose to develop improved sugarcane varieties. RIDESA was created after the extinction of the earlier governmental sugarcane breeding program Planalsucar. The creation of RIDESA was an important step towards coordinated nationwide actions for a technological support of one of the most important segments of the Brazilian economy. The consortium consists of 10 universities (UFSCar, UFRPE, UFAL, UFRRJ, UFV, UFG, UFPR, UFS, UFPI, and UFMT) that sustain and share the sugarcane flowering and crossing station denominated Serra do Ouro, in Murici/AL, and experimental units forming a national test network for sugarcane breeding. In over 20 years, RIDESA has released 78 RB sugarcane varieties, which are currently planted on 58 % of the area cultivated with sugarcane in Brazil (Daros et al. 2010).

The Federal University of São Carlos (UFSCar) is responsible for developing RB varieties for the South-Central region of Brazil, in the states of Sao Paulo and Mato Grosso do Sul. This region has the largest sugarcane area and highest cane production in Brazil. The varieties

[1] Universidade Federal de São Carlos, Centro de Ciências Agrárias, PMGCA/PAI-UFSCar, 13.600-970, Araras, SP, Brazil. *Email: monalisa@cca.ufscar.br

RB965902 and RB965917 were developed and released in 2010 by the UFSCar breeding program.

BREEDING PROGRAM

The varieties RB965902 and RB965917 are full-sibs and were obtained from a reciprocal, biparental cross between the varieties RB855536 x RB855453 (Figure 1). The crosses were carried out at the sugarcane flowering and crossing station Serra do Ouro, in Murici/AL (09° 18' S, 35° 56'W, 450 m asl). The obtained seeds were germinated and then planted in the field, establishing the first phase of selection (T1). At this stage, clones from a single clump were selected by mass selection in the first ratoon cane cycle (Breaux et al. 1963), based on criteria of important industrial morphological characteristics such as brix and stalk number (Hogarth 1987, Barding et al. 2004), flowering, pithiness and resistance to the main diseases (Matsuoka et al. 1999). The clones were compared to standard commercial varieties with early and medium/late maturity.

Based on these criteria, clones were selected which, together with early-maturing standard commercial varieties, constituted the second phase of selection (T2). In this step, the clones were established in Arraias (22° 211 S, 47° 23'W; 620 m asl) and Valparaiso (21° 13'S, 50° 52'W; 450 m asl), state of São Paulo, in an augmented block design (Federer 1956). The plots consisted of a 7-m row with one replication. The clones were evaluated in plant cane and first ratoon cane, based on the same criteria as in stage T1 together with the parameters cane weight per plot and kilo brix per plot - KBP (Kang et al. 1983).

The selected T2 clones were advanced to the third stage of selection (T3), also arranged in augmented blocks (Federer 1956). The plots consisted of two 5-m rows spaced 1.40 m apart, with two replications. The T3 genotypes were tested at 10 different locations in the Central South. The selection was based on the average performance of plant cane and first ratoon cane in different test environments. The selection criteria were similar to stage T2, plus Pol in juice % (sucrose content) and kilo pol per plot - KPP as additional parameters (Matsuoka et al. 1999).

After phase T3, the selected clones were advanced to the final experimental stage of selection, called FE. At this stage, the promising genotypes were assessed at 25 locations across the South-Central region, considering the data of four cycles (plant cane, first cane ratoon, 2nd and 3rd ratoons). The field trials were established randomized complete blocks with three replications, with early-maturing, standard commercial varieties distributed within blocks as controls. The parameters were tons of cane per hectare – TCH (cane yield), Pol in cane in % - PC (sucrose content), tons of Pol per hectare – TPH (sucrose yield) and fiber content in %. The coefficient of environmental variation, the effects of genotype-environment interaction and the clone adaptability and stability were estimated by individual (of each location) and combined analysis of variance (of all locations) (Steel and Torrie 1960). The maturation curve of the FE promising test clones was evaluated to identify the best harvest time in terms of the level of Pol in cane in % (sucrose content). The best-performing genotypes were multiplied and evaluated in the partnership units to observe the performance under production conditions (Barbosa et al. 2001, Barbosa et al. 2004). In 2010, the varieties RB965902 and RB965917 were officially released by UFSCar.

PERFORMANCE

RB965902

The growth habit of the variety is slightly decumbent and the leaves (trash) can be removed relatively easily; it has a good canopy cover and excellent ratoon regrowth from green and burnt cane, sparing an early replanting of the cane fields. The tillering capacity inboth plant and first ratoon canes is particularly high.

The fiber content is medium, maturation early to medium, flowering absent and pithiness low. In the Central South, RB965902 is indicated for harvesting between May and July. The constant PC content of this variety allows harvesting until mid-August and processing after RB855453.

The TCH of RB965902 in unfavorable environments is higher than of RB855453 and yields are more stable under improved soil and climate conditions (Figure 2). For production in the Central South, cultivation in moderate and favorable environments is recommended (Prado 2008).

With high agricultural productivity (TCH above 120 t/ha) and PC of about 13.5 %, the TPH of RB965902 is similar to the early/medium-maturing commercial standard varieties (Figure 3).

RB965917

The growth habit of the variety is upright, stalks are tall and little lodging occurs up to an age of 12 months. It has a fast canopy cover, excellent regrowth and high tillering in plant and ratoon cane. It stands out with high agricultural productivity in favorable environments and excellent performance of mechanical harvesting.

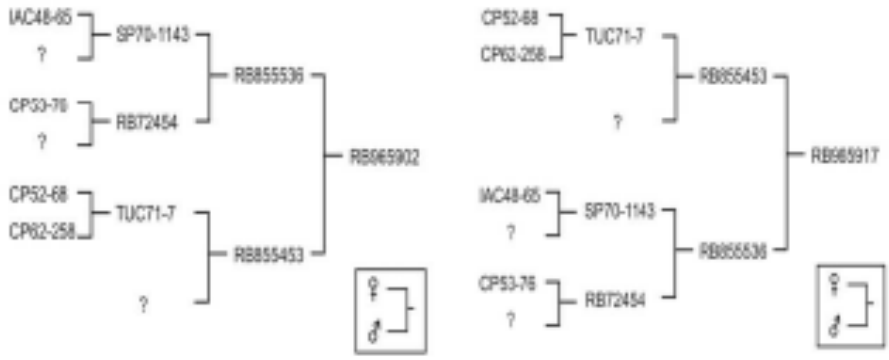


Figure 1. Pedigree of the sugarcane varieties RB965902 and RB965917.

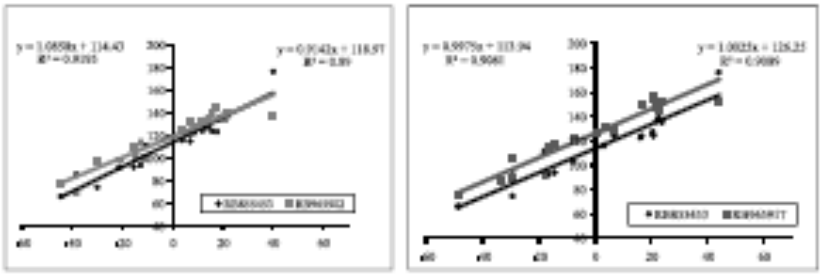


Figure 2. Adaptability and stability of the varieties RB965902 and RB965917 compared to the commercial standard RB855453. The mean data of tons of cane per hectare (TCH) in 18 field trials in a first ratoon cane cycle were adjusted based on regression analysis (Eberhart and Russell 1966).

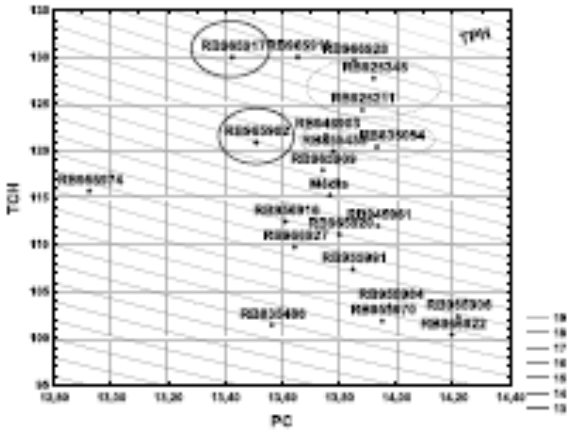


Figure 3. Isoquants of mean data of Pol in cane in % (PC) and tons of cane per hectare (TCH) in 18 field trials and 3 cycles in different production environments. The diagram shows the varieties RB965902 and RB965917 (drawn-out circles) for comparison with the standard commercial varieties (dotted circles) and clones.

The fiber content is medium, maturation early to medium, flowering absent and pithiness low. In the Central South, RB965917 is indicated for harvesting between June and August. The constant PC content of this variety allows harvesting until mid-September and can be processed after RB855453.

The variety RB965917 has higher TCH in both favorable and unfavorable environments than RB855453, with lower yield stability (Figure 2). For commercial production, RB965917 is adequate in favorable environments (Prado 2008), under similar production conditions to those of RB855453.

The high agricultural productivity (TCH above 130 t ha^{-1}) and the PC concentration of about 13.4 % of RB965917 indicate an equivalent or higher TPH than of the early/medium-maturing, commercial standard varieties (Figure 3).

OTHER FEATURES

Disease reaction

The varieties RB965902 and RB965917 in stage T3 were subjected along with other genotypes to natural disease infection and artificial inoculation tests. These tests were conducted to verify the reaction of varieties and clones against the major diseases of sugarcane in the South-Central region of Brazil.

The test was conducted in a region with high inoculum pressure, favorable for the natural infection of various diseases, such as Brown Rust (*Puccinia melanocephala*), Smut disease (*Ustilago scitaminea*), Mosaic (Sugarcane Mosaic Virus - SCM) and Leaf Scald (*Xanthomonas albilineans*). The varieties RB965902 and RB965917, as the others, were evaluated based on the number of infected tillers (infection %) over two consecutive years, in the plant and first ratoon cane cycles.

In the artificial test in a greenhouse, RB965902 and RB965917 plants were inoculated with spores of the causal fungus of Smut disease and the causal agent of Mosaic virus, according to methods described by Matsuoka (1979). The varieties were evaluated based on a grading scale for each disease, where the number of infected tillers is counted (% infection) and the genotypes are classified in resistant, intermediate and susceptible.

Based on the tests of natural infection and artificial inoculation, the varieties RB965902 and RB965917 were considered resistant to the sugarcane diseases brown rust, Smut disease, and Leaf Scald. The resistance of RB965917 was considered intermediate to Red Stripe, another important sugarcane disease.

CHARACTERIZATION BY MICROSATELLITE GENOTYPING

The molecular fingerprints of RB965917 and RB965902 were generated with a panel of 17 microsatellite markers derived from sugarcane expressed sequence tags (EST-SSRs), developed by Oliveira et al. (2007), and were compared with those of five other cultivars (RB72454, RB835486, SP80-3280, SP81-3250 and RB966928). The 17 microsatellite loci amplified a total of 136 fragments in the seven cultivars, with sizes ranging from 192 to 291 base pairs (bp).

Due to the polyploid nature of sugarcane, most of the selected EST-SSRs produced more than two alleles per genotype, on average eight alleles, ranging from 4 (SCB 64) to 12 (SCB 125) alleles. The Polymorphic Information Content (PIC) had a mean value of 0.80, ranging from 0.70 (SCB 2007) to 0.88 (SCB 40). The information of discriminatory power (DP) ranged from 1 (SCB 125, SCB and 40) and 0.91 (SCB 2007), with an average value of 0.98.

Absence and presence of fragments were coded as a binary (0, 1) matrix, which was used to generate an unweighted pair group method using arithmetic averages (UPGMA) (Figure 4) from similarity indices among the seven cultivars (Dice 1945). Results indicate that the varieties RB965902 and RB965917 were genetically closer to each other than to the others.

BASIC SEED MAINTENANCE AND DISTRIBUTION

The varieties RB965902 and RB965917 are being produced by the UFSCar Breeding Program and are available for research purposes at the Agricultural Science Center, Araras - SP, where they will be maintained for at least five years from the date of publication.

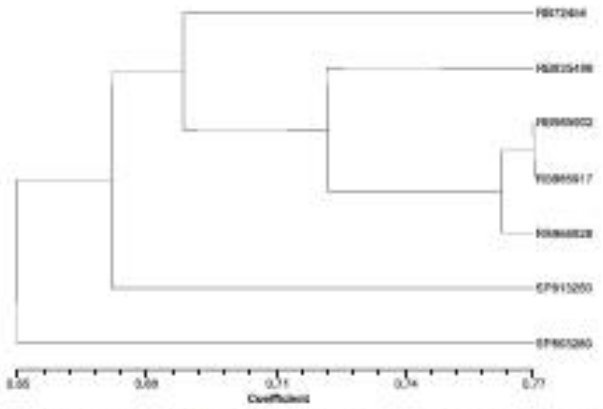


Figure 4. Dendrogram based on UPGMA representing the similarity of Dice among 7 sugarcane varieties analyzed with 17 microsatellite loci.

RB965902 e RB965917 – Variedades de cana-de-açúcar com maturação precoce/média

RESUMO - *As variedades RB965902 e RB965917 foram desenvolvidas e liberadas para colheita no início até o meio da safra (maturação precoce/média) na região Centro-Sul do Brasil. Em ambientes específicos, essas variedades superam o padrão comercial RB855453 em produção de pol por área. Apresentam resistência às principais doenças da cana-de-açúcar.*

Palavras-chave: *Saccharum*, Ridesa, programa de melhoramento de cana.

REFERENCES

Barbosa MHP, Silveira LCI, Oliveira MW, Souza VFM and Ribeiro SNN (2001) RB867515 Sugarcane cultivar. **Crop Breeding and Applied Biotechnology 1**: 437-438.

Barbosa MHP, Silveira LCI, Souza VFM and Ribeiro SNN (2004) RB928064 - Sugarcane cultivar. **Crop Breeding and Applied Biotechnology 4**: 356-359.

Berding N, Hogarth M and Cox M (2004) Plant improvement of sugarcane. In James GI. (ed.) **Sugarcane**. Blackwell Science, Oxford, p. 1-19.

Breaux RD, Hebert LP and Fanguy HP (1963) Defects for which sugarcane seedlings are eliminated at the U.S. Sugar Cane Field Station, Houma, Louisiana. In **Proceedings of Congress of International Society of Sugarcane Technologists**. Elsevier, Amsterdam, p. 421-424.

Daros E, Zambon JLC, Oliveira RA and Bespalhok Filho JC (eds.) (2010) **Liberação nacional de novas variedades "RB" de cana-de-açúcar**. AJIR, Curitiba, 64p.

Dice LR (1945) Measures of the amount of ecological association between species. **Ecology 26**: 297-307.

Eberhart SA and Russell WA (1966) Stability parameters for comparing varieties. **Crop Science 6**: 36-40.

Federer WT (1956) Augmented (or Hoonuiaku) designs. Hawaiian Planters' Record 55: 191-208.

Hogarth DM (1987) Genetics of sugarcane. In Heinz DJ (ed.) **Sugarcane Improvement through breeding**. Elsevier, Amsterdam, p. 255-271.

Kang MS, Miller JD and Tai PYP (1983) Genetic and phenotypic path analysis and heritability in sugarcane. **Crop Science 23**: 643-647.

Matsuoka S (1979) Método para pré-testagem de clones de cana-de-açúcar ao carvão e ao mosaico conjuntamente. In **I Congresso nacional da sociedade dos técnicos açucareiros e alcooleiros do Brasil**. STAB, Maceió, p. 231-233.

Matsuoka S, Garcia AAF and Arizono H (1999) Melhoramento da cana-de-açúcar. In Bordin A (ed.) **Melhoramento de espécies cultivadas**. Editora UFV, Viçosa, p. 205-252.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF and Souza AP (2007) Characterization of new polymorphic functional markers for sugarcane. **Genome 52**: 191-209.

Prado H (2008) **Pedologia fácil: aplicações na agricultura**. Hélio do Prado, Piracicaba, 45p.

Sreenivasan TV, Ahloowalia BS and Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed.) **Sugarcane Improvement through breeding**. Elsevier, Amsterdam, p. 211–253.

Steel RGD and Torrie JH (1960) **Principles and procedures of statistics**. McGraw-Hill, New York, 481p.

Crop Breeding and Applied Biotechnology 15: 193-196, 2015
Brazilian Society of Plant Breeding. Printed in Brazil

CULTIVAR RELEASE

http://dx.doi.org/10.1590/1984-70332015v15n3c34

RB975952 – Early maturing sugarcane cultivar

Monalisa Sampaio Carneiro[1*], Roberto Giacomini Chapola[1], Antônio Ribeiro Fernandes Júnior[1], Danilo Eduardo Cursi[1], Fernanda Zatti Barreto[1], Thiago Willian Almeida Balsalobre[1] and Hermann Paulo Hoffmann[1]

Received 29 July 2014

Accepted 7 November 2014

Abstract – *RB975952 is an early maturing sugarcane cultivar released for the South-Central region of Brazil. It should be harvested between April and May, and it is recommended for planting in environments with medium to high production potential. RB975952 has high resistance levels to the main diseases of the crop, it also has a good shoot development after mechanical harvesting, and high sucrose yields.*

Key words: Saccharum spp., *selection, improvement.*

INTRODUCTION

The Genetic Improvement Program of the Federal University of Sao Carlos – PMGCA/UFSCar (www.pmgca.ufscar.br) is part of the Inter University Network for the Development of Sugar and Energy Sector – RIDESA (www.ridesa.com.br), a network of 10 public Federal Universities with the purpose to develop improved sugarcane cultivars. PMGCA/UFSCar is responsible for developing RB cultivars for the South-Central region of Brazil, in the states of Sao Paulo and Mato Grosso do Sul. This region has the largest sugarcane area, and the highest sugarcane production in Brazil. The process for release of new cultivars of sugarcane is performed in several locations and crop years (Mattos et al. 2013). The development of cultivars with an early maturity cycle is one of the main objectives of PMGCA/UFSCar.

BREEDING PROGRAM

RB975952 cultivar was obtained from a biparental cross between RB835486 and RB825548 (Figure 1). The cross was carried out at the sugarcane flowering and crossing station Serra do Ouro, *in Murici, state of Alagoas* (lat 09° 18' S, long 35° 56' W, alt 450 m asl). The obtained seeds were germinated and then planted in the field, establishing the first selection stage (T1). At this stage, clones from a single clump were selected by mass selection in the first sugarcane ratoon cycle (Breaux et al. 1963), based on criteria of important industrial and morphological characteristics, such as brix and stalk number (Hogarth 1987, Berding et al. 2004), flowering, pithiness and resistance to the main diseases (Matsuoka et al. 1999). Clones were compared to standard commercial varieties with early and medium/late maturity.

Clones selected in T1 with brix equal to or higher than early-maturing standard varieties constituted the second selection stage (T2), together with early-maturing standard varieties. In this stage, clones were established in Araras (lat 22° 21' S, long 47° 23' W, alt 620 m asl) and Valparaiso (lat 21° 13' S, long 50° 52' W, alt 450 m asl), state of Sao Paulo, in an augmented block design (Federer 1956). Plots

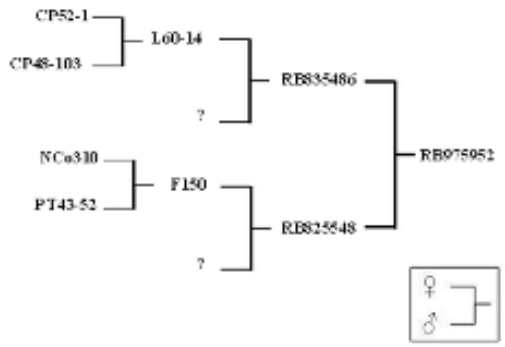


Figure 1. Pedigree of RB975952 sugarcane cultivar.

[1] Universidade Federal de São Carlos, Departamento de Biotecnologia e Produção Vegetal e Animal, Laboratório de Biotecnologia de Plantas, Araras, São Paulo, 13.600-970, Brazil. *E-mail: monalisa@cca.ufscar.br

consisted of a 7-m row with no replication. Clones were evaluated in plant-cane, and in first and second ratoon cycles, based on the same criteria as in stage T1, together with the parameters stalk weight per plot and kilogram of brix per plot – KBP (Kang et al. 1983). The third selection stage (T3) and final experimental stage of selection (FE) were carried out according to Carneiro et al. (2011). The variables evaluated were cane yield (TCH), sucrose content (PC in %), tons of pol per hectare (TPH - sucrose yield), and fiber content (%). The coefficient of environmental variation, the effects of genotype-environment interaction, and clone adaptability and stability were estimated by individual (of each location) and combined analysis of variance (of all locations) (Steel and Torrie 1960). The maturation curve of the FE promising clones was evaluated to identify the best harvest time in terms of PC% level. The best-performing genotypes were multiplied and evaluated in the partnership units to observe the performance under production conditions (Barbosa et al. 2001, Barbosa et al. 2004, Melo et al. 2014).

PERFORMANCE

RB975952

The growth habit of this cultivar is slightly decumbent, and leaves (trash) can be easily removed; it has good canopy cover and excellent ratoon regrowth from green and burnt sugarcane, sparing an early replanting of sugarcane fields. Tillering capacity in both plant and ratoon cane is good. RB975952 has medium fiber content, early maturation, rare flowering and low pithiness. In the South-Central region, RB975952 is indicated for harvesting between April and May (Figure 2). The constant sucrose content of this variety allows harvesting until mid-June. TCH of RB975952 in unfavorable environments is higher than that of RB855453, and yields are more stable under favorable conditions (Figure 3). For commercial production, RB975952 is recommended for planting in environments of medium to high fertility. The high agricultural productivity (TCH above 114.9 t ha^{-1}) and sucrose content of about 14.5% of RB975952 indicate an equivalent or higher TPH than of the early-maturing commercial standard varieties (Figure 3). Experiments were carried out at 13 different locations in São Paulo state during three harvests.

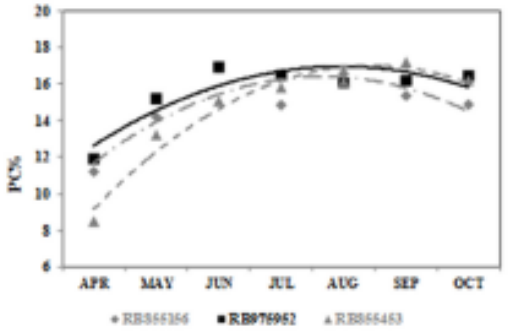


Figure 2. Maturation curve of RB975952 cultivar, compared to RB855156 and RB855453 standard commercial cultivars, for sucrose content (%) in cane.

OTHER FEATURES

Disease reaction

RB975952 was subjected to natural disease infection and artificial inoculation tests, along with other genotypes. These tests are carried out to verify the reaction of varieties and clones regarding the major diseases of sugarcane in the South-Central region of Brazil (Table 1). Tests were carried out in regions with high inoculum pressure, favorable to

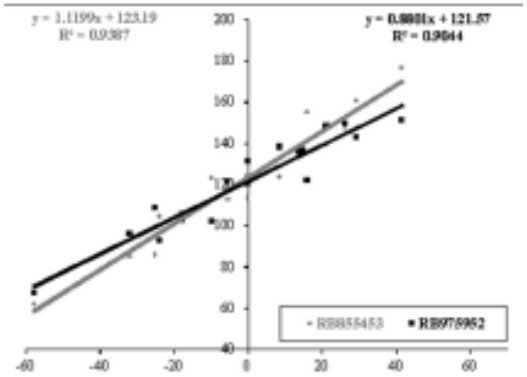


Figure 3. Adaptability and stability of RB975952 cultivar, compared to RB855453 standard commercial cultivar. The average data of sugar yield (TCH) in 17 field trials in first ratoon cane cycle were adjusted based on regression analysis.

Table 1. Disease reactions and presence (+) or absence (-) of the *Bru1* gene in RB975952 and RB855453 sugarcane cultivars, in South-Central of Brazil

Disease	Cultivar RB975952	Cultivar RB855453
Smut	R	R
Brown rust	R	R
Bru1	+	+
Orange rust	R	R
Mosaic	R	R
Leaf Scald	R	R

R= resistant
+ = presence of Bru1 molecular marker

natural infection of various diseases, such as brown rust (*Puccinia melanocephala*), orange rust (*P. kuehnii*), smut (*Sporisorium scitamineum*), mosaic (Sugarcane Mosaic Virus - SCMV), and leaf scald (*Xanthomonas albilineans*). RB975952, as the others, was evaluated based on the number of infected tillers (infection %) for smut, mosaic and leaf scald, and based on the leaf area with symptoms for brown rust and orange rust (Amorim et al. 1987, Klosowski et al. 2013). Taking into account the presence of the *Bru1* gene, and the lack of naturally infected plants (rating = 0), RB975952 was determined to be resistant to brown rust.

In the greenhouse artificial test, RB975952 plants were inoculated with the causal agents of smut and mosaic, according to the methods described by Matsuoka (1979). Cultivars were evaluated based on a rating scale for each disease, where the number of infected tillers is counted (% infection), and the genotypes are classified as resistant, intermediate and susceptible. Based on natural infection and artificial inoculation tests, RB975952 was considered resistant to brown rust, orange rust, smut, mosaic and leaf scald.

Characterization by microsatellite genotyping and Bru1 marker

SSR markers used to molecular fingerprints of RB975952 were generated with a panel of 383 microsatellite markers derived from sugarcane expressed sequence tags (EST-SSRs), developed by Oliveira et al. (2009) and Marconi et al. (2011), and were compared with those of eight other cultivars (RB925211, RB835054, RB855453, SP91-1049, RB835486, RB855156, RB825548 and RB966928). The 27 EST-SSR loci amplifications revealed high levels of polymorphism among the nine sugarcane

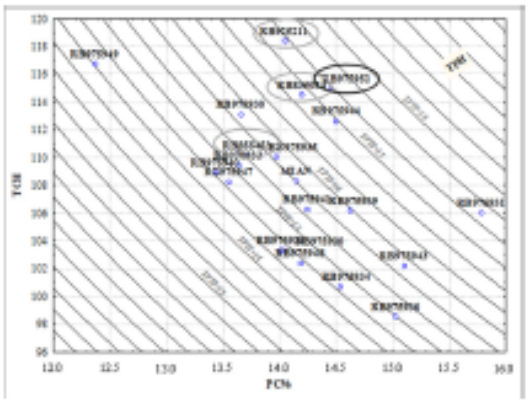


Figure 4. Isoquants of average data of sucrose content (PC%), and cane yield (TCH) in 13 field trials and three cycles in different production environments. The diagram shows RB975952 cultivar (black circle) for comparison with standard commercial cultivars (gray circles) and clones.

genotypes, and detected 360 alleles polymorphic, with a range of 5 (ESTC19) to 23 (ESTB 432), and sizes ranging from 142 to 278 base pairs (bp). The Polymorphic Information Content (PIC) value was calculated by Pinto et al. (2004) and had an average value of 0.84, ranging from 0.92 (ESTB 423) to 0.50 (ESTB99). The information of discriminatory power (DP) was calculated based on Tessier et al. (1999), and ranged from 1 and 0.89, with an average value of 0.99. Polymorphic bands were used to construct a binary matrix to evaluate the genetic similarity among all the genotypes (Santos et al. 2014). The EST-SSR-based genetic similarity (SSR-GS) among all of the genotypes was estimated according to the Simple Matching similarity coefficient. The corresponding genetic similarity matrix was used to generate a dendrogram based on the Unweighted Pair Group Method with the Arithmetic Average

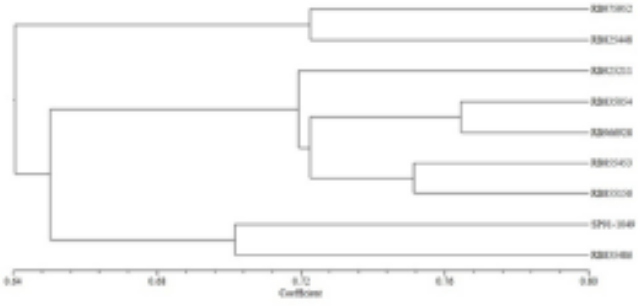


Figure 5. Dendrogram of nine sugarcane cultivars revealed by UPGMA cluster analysis of SSR genetic similarity (Simple Matching's coefficient) estimates, using 360 SSR polymorphic bands obtained by 27 primer combinations.

(UPGMA) algorithm (Figure 5). All analyses were carried out using NTSYSpc 2.11X (Rohlf 2000). Results indicate that the genetic similarities based on the Simple Matching coefficient varied from 0.58 to 0.76 with RB966928 and RB835054, meaning that these cultivars were genetically closer to each other than to the others.

REFERENCES

Amorim L, Bergamin Filho A, Sanguino A, Cardoso COM, Moraes VA and Fernandes CR (1987) Metodologia de avaliação de ferrugem da cana-de-açúcar (*Puccinia melanocephala*). **Boletim Técnico Copersucar 39**: 13-16.

Barbosa MHP, Silveira LCI, Oliveira MW, Souza VFM and Ribeiro SNN (2001) RB867515 Sugarcane cultivar. **Crop Breeding and Applied Biotechnology 1**: 437-438.

Barbosa MHP, Silveira LCI, Souza VFM and Ribeiro SNN (2004) RB928064 - Sugarcane cultivar. **Crop Breeding and Applied Biotechnology 4**: 356-359.

Berding N, Hogarth M and Cox M (2004) Plant improvement of sugarcane. In James GL (Ed) **Sugarcane**. Blackwell Science, Oxford, p. 1-19.

Breaux RD, Hebert LP and Fanguy HP (1963) Defects for which sugarcane seedlings are eliminated at the U.S. Sugar Cane Field Station, Houma, Louisiana. In **Proceedings of congress of international society of sugarcane technologists**. Elsevier, Amsterdam, p. 421-424.

Carneiro MS, Rosa JRBF, Barreto FZ, Balsalobre TWA, Chapola RG, Vieira MAS, Bassinello AI, Hoffmann HP (2011) RB965902 and RB965917 – Early/medium maturing sugarcane varieties. **Crop Breeding and Applied Biotechnology** II: 280-285.

Federer WT (1956) Augmented (or Hoonuiaku) designs. **Hawaiian Planters' Record 55**: 191-208.

Hogarth DM (1987) Genetics of sugarcane. In Heinz DJ (Ed) **Sugarcane improvement through breeding**. Elsevier, Amsterdam, p. 255-271.

Kang MS, Miller JD and Tai PYP (1983) Genetic and phenotypic path analysis and heritability in sugarcane. **Crop Science 23**: 643-647.

Klosowski AC, Ruaro L, Bespalhok Filho JC and De Mio LLM (2013) Proposta e validação de escala para a ferrugem alaranjada da cana-de-açúcar. **Tropical Plant Pathology 38**: 166-171.

Marconi TG, Costa EA, Miranda HR, Mancini MC, Cardoso-Silva CB, Oliveira KM and Souza AP (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. **BMC Research Notes 4**: 264.

Mattos PHC, Oliveira RA, Filho JCB, Daros E and Veríssimo MAA (2013) Evaluation of sugarcane genotypes and production environments in Paraná by GGE biplot and AMMI analysis. **Crop Breeding and Applied Biotechnology 13**: 83-90.

Matsuoka S (1979) Método para pré-testagem de clones de cana-de-açúcar ao carvão e ao mosaico conjuntamente. In **I Congresso nacional da sociedade dos técnicos açucareiros e alcooleiros do Brasil**. STAB, Maceió, p. 231-233.

Matsuoka S, Garcia AAF and Arizono H (1999) Melhoramento da cana-de-açúcar. In Borém A (Ed) **Melhoramento de espécies cultivadas**. Editora UFV, Viçosa, p. 205-252.

Melo LJOT, Daros E, Neto DES, Chaves A, Silva LJ, Silva AEP and Melo TTAT (2014) CULTIVAR RELEASE - RB962962, A Sugarcane Cultivar For Late Harvest. **Crop Breeding and Applied Biotechnology 14**: 132-135.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF and Souza AP (2009) Characterization of new polymorphic functional markers for sugarcane. **Genome 52**: 191-209.

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, and Souza AP (2004) Survey in the expressed sequence tag database (SUCEST) for simple sequence repeats. **Genome 47**: 795-804.

Rohlf FJ (2000) NTSYSpc: numerical taxonomy and multivariate analysis system, version 2.11X. Applied Biostatistics, New York.

Santos JM, Barbosa GVS, Neto CER and Almeida C (2014) Efficiency of biparental crossing in sugarcane analyzed by SSR markers. **Crop Breeding and Applied Biotechnology 14**: 102-107.

Steel RGD and Torrie JH (1960) **Principles and procedures of statistics**. McGraw-Hill, New York, 481p.

Tessier C, David J, This P, Boursiquot JM and Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. **Theoretical and Applied Genetics 98**: 171-177.

BASIC SEED MAINTENANCE AND DISTRIBUTION

RB975952 has been produced by PMGCA/UFSCar and is available for research purposes at the Agricultural Science Center (CCA/UFSCar), in Araras, state of Sao Paulo, where it will be maintained for at least five years from the date of publication.

UNICAMP
COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br
IB

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "***MAPEAMENTO DE QTLs EM POPULAÇÃO DERIVADA DE CRUZAMENTO COMERCIAL BI-PARENTAL EM CANA-DE-AÇÚCAR***", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: Thiago Willian Almeida Balsalobre
Nome do(a) aluno(a): Thiago Willian Almeida Balsalobre

Assinatura: _____
Nome do(a) orientador(a): Anete Pereira de Souza

Data: 25/10/2016

Profa. Dra. Rachel Meneguello
Presidente
Comissão Central de Pós-Graduação
Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **MAPEAMENTO DE QTLs EM POPULAÇÃO DERIVADA DE CRUZAMENTO COMERCIAL BI-PARENTAL EM CANA-DE-AÇÚCAR**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 25 de outubro de 2016

Assinatura : Thiago Willian Almeida Balsalobre
Nome do(a) autor(a): **Thiago Willian Almeida Balsalobre**
RG n.° 40204589-0

Assinatura : _____
Nome do(a) orientador(a): **Anete Pereira de Souza**
RG n.° 8680325